LINGUÍSTICA PARA O ENSINO SUPERIOR 2

# PROSÓDIA

PLÍNIO A. BARBOSA

711
P

parábola

Scanned with CamScanner

Direção: Andréia Custódio
Editores científicos: Tommaso Raso [UFMG]
Celso Ferrarezi Jr. [UNIFAL-MG]
Diagramação e capa: Telma Custódio
Revisão: Thiago Zilio Passerini

Clas. G 469.07 J1
Cult. B238p
v.1
Tombo 165753

DOAÇÃO
DE Mauro Damaceno
EM 25/11/25
VALOR EST. R$ 40,00

**CIP-BRASIL. CATALOGAÇÃO NA FONTE**
**SINDICATO NACIONAL DOS EDITORES DE LIVROS, RJ**

B211p

Barbosa, Plínio A.
Prosódia / Plínio A. Barbosa ; [coordenação Tommaso Raso, Celso Ferrarezi Jr]. - 1. ed. - São Paulo : Parábola, 2019.
136 p. ; 23 cm. (Linguística para o ensino superior ; 2)

Inclui bibliografia
ISBN 978-85-7934-163-2

1. Língua portuguesa - Estudo e ensino (Superior). 2. Língua portuguesa - Prosódia. I. Raso, Tommaso. II. Ferrarezi Jr., Celso. III. Título. IV. Série.

19-56334

CDD: 469.16
CDU: 811.134.3:801.6

Leandra Felix da Cruz - Bibliotecária - CRB-7/6135

Direitos reservados à
**PARÁBOLA EDITORIAL**
Rua Dr. Mário Vicente, 394 - Ipiranga
04270-000 São Paulo, SP
pabx: [11] 5061-9262 | 5061-8075 | fax: [11] 2589-9263
home page: www.parabolaeditorial.com.br
e-mail: parabola@parabolaeditorial.com.br

Todos os direitos reservados. Nenhuma parte desta obra pode ser reproduzida ou transmitida por qualquer forma e/ou quaisquer meios (eletrônico ou mecânico, incluindo fotocópia e gravação) ou arquivada em qualquer sistema ou banco de dados sem permissão por escrito da Parábola Editorial Ltda.

**ISBN: 978-85-7934-163-2**

1ª edição, 1ª reimpressão: março de 2020

© do texto: **Plínio A. Barbosa, 2019.**
© da edição: **Parábola Editorial, São Paulo, 2019.**

Scanned with CamScanner

# Agradecimentos

A Tommaso Raso e à equipe de editores pela oportunidade e sugestões.

Minha gratidão pelas tantas interações durante e fora das aulas com alunos e colegas é imensa, pois eles moldaram a forma deste livro, especialmente meus orientandos.

Diversas colaborações de pesquisa me fizeram refletir sobre vários dos temas apresentados. Em especial destaco a interlocução com Sandra Madureira, Zuleica Camargo, Oliver Niebuhr, Philippe Boula de Mareüil, Anders Eriksson, Hansjörg Mixdorff, Tommaso Raso, Heliana Mello, Catarina Oliveira e António Teixeira. A eles, meu muito obrigado.

Gratidão também a meus colegas de Instituto com os quais discuti sobre prosódia, especialmente Bernadete Abaurre, Pablo Faria, Lívia Oushiro e Filomena Sandalo. Aos funcionários do Instituto de Estudos da Linguagem pelo apoio de sempre quanto a ensino, pesquisa e extensão. Destaco especialmente Cláudio, Rose, Diurielson e Emerson.

Às agências de pesquisa CAPES, CNPq e FAPESP, que tornaram os trabalhos, a participação e organização de eventos sobre a prosódia da fala possíveis.

A manutenção da integridade afetiva, psíquica e espiritual foi fundamental para fazer um trabalho que pudesse alegrar todos a meu lado, pois creio ser essa uma das funções do trabalho acadêmico. Sem o apoio da família, em todos os aspectos e momentos, isso não seria possível. A ela agradeço imensamente, reconhecendo um débito que não pode ser pago.

A minha esposa Rose (Rosinha), pelo apoio, paciência diante de tantas horas longe dela por conta da pesquisa e do ensino. Por sua radiante alegria, expressiva como a prosódia, deixo aqui meu sinal de amor e gratidão.

Aos nossos filhos, Denis, Breno, Lucas e Aulísio, que participaram seja lendo uma parte deste trabalho, seja como voluntários de experimentos, seja me incentivando nesta obra, minha gratidão profunda. Além deles, os demais filhos, afilhados e amigos tornaram nossa casa um oásis de alegria que aliviou as dificuldades diárias do trabalho. Muito obrigado, queridos Lisinho, Ariane, Éder (Edinho), Amaurys, Filipe, Jonesmar (John), Tháles, Marília, Murilo (Lillo), Ticiana (Tici), Vinícius (Vini), Giovanna (Gi), Paulo Henrique, Carol, Matheusinho, Leonardo, Gabriel, César, Flávio, Tiago e Emerson.

Aos meus irmãos Cláudio, Erik, Gerson, Hans, Lippe, Luciano, Luiz Carlos, Olli e Ricardo.

Aos meus pais e aos meus sogros, pela vida e educação minha e de minha Rose.

A Deus, *sine Quo nihil esse posset*.



Scanned with CamScanner

# Sumário

**Primeiras palavras** ..... 11

**Apresentação** ..... 13

**Anacruse** ..... 17

CAPÍTULO 1 – **Introdução** ..... 19
1.1 Correlatos físicos da prosódia ..... 21
1.2 Correlatos perceptivos da prosódia ..... 30
1.3 Softwares para o estudo da prosódia ..... 34

CAPÍTULO 2 – **Unidades prosódicas** ..... 37
2.1 A sílaba ..... 37
2.2 O pé ..... 41
2.3 A palavra fonológica ou prosódica ..... 43
2.4 O grupo acentual (*stress group*) ..... 43
2.5 O sintagma entoacional (*intonational phrase*) ..... 45
2.6 O enunciado (entoacional) ..... 47

CAPÍTULO 3 – **O ritmo da fala** ..... 49
3.1 Línguas de ritmo silábico e de ritmo acentual ..... 50
3.2 Funções rítmicas de segmentação e proeminência ..... 51
3.3 Taxa de elocução (*speech rate*) e de articulação (*articulation rate*) ..... 53
3.4 Procedimentos experimentais da pesquisa sobre ritmo da fala ..... 55

CAPÍTULO 4 – **Entoação da fala** ..... 67
4.1 Tom de fronteira ..... 68
4.2 Acento de pitch ..... 71
4.3 Alinhamento tonal ..... 71



Scanned with CamScanner

4.4 Aspectos entoacionais de longo termo ..... 72
4.5 Procedimentos experimentais da pesquisa sobre entoação da fala ..... 75

CAPÍTULO 5 – **A prosódia da fala e suas interfaces** ..... 81
5.1 Prosódia e sintaxe ..... 82
5.2 Prosódia e significado ..... 84
5.3 Prosódia e discurso ..... 87
5.4 Prosódia e processamento ..... 90
5.5 Disprosódia ..... 93
5.6 Prosódia e expressividade ..... 96
5.7 Prosódia e indivíduo ..... 104

CAPÍTULO 6 – **Teorias e modelos** ..... 109
6.1 Teorias prosódicas pautadas em princípios biológicos ..... 109
6.2 Modelos de geração da melodia da fala ..... 111
6.3 Modelos de geração da duração e do ritmo da fala ..... 114
6.4 O papel da prosódia em modelos de percepção da fala ..... 116

CAPÍTULO 7 – **Para saber mais** ..... 119
7.1 Capítulo 1 ..... 120
7.2 Capítulo 2 ..... 123
7.3 Capítulo 3 ..... 123
7.4 Capítulo 4 ..... 125
7.5 Capítulo 5 ..... 125
7.6 Capítulo 6 ..... 126

**Referências bibliográficas** ..... 127

# Primeiras palavras

Uma das principais missões da Associação Brasileira de Linguística [ABRALIN] é divulgar o conhecimento da área em camadas cada vez mais amplas da população, tentando se fazer presente na sociedade como agente de educação completo, não apenas como órgão de debates acadêmicos internos.

Essa missão tem se realizado pelo recurso a instrumentos e canais variados, mediante iniciativas de diferentes naturezas, entre as quais a promoção de redes internacionais, a produção de eventos em sedes tradicionalmente afastadas dos principais centros acadêmicos, o uso de mídia alternativa ao texto escrito para popularizar as diferentes disciplinas da área e um empenho social articulado com outras instâncias acadêmico-sociais.

A *Coleção Linguística para o ensino superior*, da Parábola Editorial, editada por Tommaso Raso e Celso Ferrarezi, apresenta-se como uma iniciativa ampla, especificamente pensada para oferecer aos alunos do ensino superior um quadro completo das disciplinas linguísticas, com um olhar para o panorama internacional da área e meios de divulgação adequados aos nossos tempos. Por essa razão, a coleção tem a chancela da ABRALIN.

Faltava ao panorama editorial e educacional do país um instrumento como este, capaz de aproximar ainda mais o Brasil do padrão vigente em meios universitários internacionais. A ABRALIN, sempre atenta aos meios necessários para o fortalecimento da área no País, sem nunca perder de vista o compromisso com a qualidade, apoia e parabeniza a iniciativa da Parábola Editorial.

**Miguel Oliveira Jr.**
Presidente da ABRALIN



Scanned with CamScanner

# Apresentação

O objetivo da coleção *Linguística para o ensino superior* é o de fornecer ao Brasil livros de referência, escritos em nossa língua e por autores consagrados, como instrumentos de estudo em todos os cursos que demandem conhecimentos de base nessa ampla área de estudos que é a linguística. O principal público-alvo é, portanto, constituído pelos alunos universitários, mas leitores interessados nos estudos linguísticos poderão encontrar, em todas as obras da coleção, textos, ao mesmo tempo, bem fundamentados e amigáveis. A coleção foi idealizada de maneira integrada e coerente, com coordenadas que se repetem em todos os volumes. Isso uniformiza, tanto quanto possível, as obras e facilita o contato dos leitores com a coleção.

Uma das intenções dos organizadores e do editor foi a de olhar para a rica manualística existente no âmbito internacional — em primeiro lugar, a produção em língua inglesa, mas também nas línguas alemã, francesa e italiana —, procurando estabelecer contatos entre a produção brasileira e a internacional. Primamos para que cada volume fosse escrito por um único autor, de forma que a exposição fosse a mais orgânica possível e a coerência informativa e didática envolvesse não somente o momento da organização da obra, mas também o da execução intelectual e material.

Por essas razões todas, os volumes foram elaborados de forma a oferecer o que de mais consolidado e consensual existe em linguística, sem direcionar o conteúdo exposto para vertentes teó-



Scanned with CamScanner

ricas específicas. Cremos que a escolha de um arcabouço teórico específico (gerativismo, cognitivismo, estruturalismo, otimalismo ou arcabouços específicos de matriz funcionalista) seja um passo que só deve ser dado pelos alunos depois de eles construírem uma sólida base conceitual e metodológica na disciplina e de terem tomado conhecimento da variedade e da riqueza do panorama teórico disponível em cada disciplina e das motivações dessa variedade. Dessa forma, no caso dos volumes desta coleção, os conceitos e as explicações sobre a área indicada no título são aqueles que se consolidaram ao longo da linguística moderna como sendo os mais aceitos nas diversas vertentes, sempre que isso for possível, e, quando oportuno, mostrando possíveis abordagens diferentes às questões tratadas.

Como editores científicos da coleção, cremos que organizar e disponibilizar um sólido e abrangente conjunto de conhecimentos de referência não é tolher os horizontes dos estudantes ou eliminar o objetivo das disciplinas. Ao contrário disso, é proporcionar-lhes uma formação de base aberta e livre, oferecendo-lhes formas de avanço mais sólidas e mais cientificamente produtivas.

Com esse mesmo objetivo, os volumes apresentam sempre uma seção final, intitulada "Para saber mais". Nela, o leitor encontrará extensas indicações para o aprofundamento bibliográfico, organizadas de forma temática e com comentários para cada obra indicada. São apresentadas obras de referência em português e em outras línguas, com um duplo objetivo: por um lado, introduzir os leitores aos clássicos e às principais novidades do panorama internacional; por outro, estimulá-los ao aprofundamento conceitual e metodológico.

Mas, as obras de nossa coleção não terminam nos volumes físicos. A Parábola manterá uma página na internet associada a cada volume para abrigar materiais complementares escolhidos pelos autores. Nela, estarão presentes exercícios e, frequentemente, materiais adicionais e de natureza multimidiática: textos de aprofundamento, bibliografia suplementar, arquivos de áudio e vídeo, imagens e outros formatos compatíveis com os objetivos da coleção. Essas páginas serão alimentadas periodicamente, mantendo o trabalho vivo ao longo do tempo. No caso em que aprofundamentos já tenham sido previstos desde a elaboração do volume, o símbolo avisa o leitor de que há mais material no *site* para o tema em estudo.

Todo esse esforço, que envolveu uma grande equipe de trabalho entre autores, organizadores e a Parábola, é uma tentativa de fornecer a alunos e docentes um material consistente, e, a despeito de todas as dificuldades, de ampliar os horizontes da moderna *linguística* no Brasil.

Desejamos a todos uma excelente leitura!

**Tommaso RASO [UFMG]**
**Celso FERRAREZI Jr. (UNIFAL-MG)**



Scanned with CamScanner

# Anacruse

Este livro é fruto de mais de 25 anos de pesquisa sobre a prosódia, iniciada em 1990 durante o doutorado sobre o ritmo do francês e continuada no exercício de pesquisa, docência e extensão majoritariamente com estudos sobre o português brasileiro e europeu.

A pesquisa é primordial para um trabalho sempre atualizado que interpele os colegas e alunos nos vários momentos de troca de conhecimento, mas um livro voltado para a graduação é fruto primordialmente de uma preocupação com a aquisição do conhecimento por um(a) aluno(a) que está tendo contato pela primeira vez com o assunto. Procurei fazer com que pessoas de variadas áreas, mesmo de fora da Linguística, lessem versões iniciais deste livro. Suas observações, ressalvados meus eventuais erros de execução, contribuíram para a maneira como escrevi, que inclui vários exemplos. Esses exemplos são acompanhados de material de apoio no *site,* todos assinalados pelo símbolo " ".

Desejo que a leitura seja agradável e que a incursão por terrenos prosódicos seja motivo de alegria e gere uma vontade de aprofundamento na pesquisa experimental nessa área. Como docente, estou sempre à disposição de todos os que me procuram. Boa leitura e bons estudos.

**Plínio A. Barbosa**
Departamento de Linguística
Instituto de Estudos da Linguagem
Unicamp

Campinas, 30 de março de 2019



Scanned with CamScanner

CAPÍTULO 1

# Introdução

Aprendendo a prosódia de uma língua se entra
mais intimamente no espírito da nação que a fala mais
do que por qualquer outro tipo de estudo que possa haver.
[Madame de Staël, "Sobre a Alemanha"]

Que mistério é esse, evocado pela romancista suíça, que associa a **prosódia** ao espírito de uma nação, à intimidade com uma língua? Faz-nos pensar ser a prosódia um modo intrínseco de expressão ou do "modo" de falar de uma comunidade. De fato, essa interpretação não está distante da acepção da palavra "prosódia" na fonética.

De origem grega, o termo *prosódia* (προσωδία), em que προσ significa "em direção a" ou "junto com" e ωδία significa "canto", foi usado pela primeira vez no livro *República* de Platão na expressão *phthongous te kai prosódias* (algo como "segmentos *vs.* prosódias") para opor aquilo que é dito (*phthongous*) à forma como é dito (*prosōdias*). A expressão aparece no contexto de uma discussão sobre a oposição entre narrativa pura e imitação em que Sócrates sugere uma censura à imitação, especialmente quando são calques do modo de falar de determinados heróis gregos, por essa imitação poder ferir a reputação desses heróis (Liddel, Scott e Jones, 1996).

Da expressão acima se deduz que Platão se refere ao termo "prosódia" em contraste com o conteúdo dos enunciados, considerando-a um componente sonoro em "sintonia com o canto". Portanto, desde a origem, "prosódia" retém ao menos a acepção de "modo de falar", seja ele intencional ou não.



Scanned with CamScanner

Com o passar dos séculos, no entanto, o termo agregará sentidos especializados associados a outros termos muito usados nas áreas de fonética, fonologia e ciências da fala como acento, entoação[1], ênfase e ritmo, bem como sentidos associados ao discurso, a atitudes e emoções, todos eles imbricados com fatores sociais e biológicos como gênero, identidade, classe social, entre outros.

Aos estudos de prosódia cabe a análise fonética e fonológica das relações entre unidades silábicas, que são a base de constituição de relações entre unidades superiores, no intuito de moldar um modo de falar para determinado fim. Assim, o estudo da prosódia não considera diretamente o conteúdo segmental, ou o "que se diz", e sim a forma sonora e sua função ligadas ao "como se diz".

Para empreender esse estudo, a análise prosódica se dá em eixos linguísticos tradicionais, o *eixo sintagmático* e o *eixo paradigmático*. O primeiro eixo diz respeito à relação entre elementos presentes no trecho falado, enquanto o segundo diz respeito à relação entre trechos em contraste de enunciados distintos.

Como exemplo de análise prosódica paradigmática, consideremos o contraste entre dois atos de fala, uma asserção e uma questão, para a mesma sequência de palavras. Assim, o enunciado interrogativo 1.1 correspondente à sentença "Pedro fez bem seu trabalho?" pode ser comparado, via análise, ao enunciado assertivo 1.2 correspondente à sentença "Pedro fez bem seu trabalho". A escuta atenta dos dois enunciados revela que a principal diferença reside na emissão mais aguda da sílaba tônica final ("ba") no enunciado interrogativo.

Como exemplo de análise prosódica sintagmática, consideremos o enunciado 1.3 correspondente à sentença "João comprou a moto VERDE.", em que a última palavra é realizada por seu enunciador com *foco contrastivo*, isto é, com destaque em relação às demais palavras, para deixar claro a seu interlocutor que a moto a que se refere não é, por exemplo, a moto vermelha também à sua frente. A análise sintagmática pode, nesse exemplo, descrever os aspectos sonoros da palavra "verde" em relação aos da palavra precedente ("moto"), ambas enunciadas na mesma frase. A palavra focada é pronunciada de maneira mais forte, mais aguda e mais lenta do que a palavra "moto".

A importância dos estudos da prosódia para os estudos linguísticos emergirá claramente para o leitor a partir dos próximos capítulos, mas podemos dar dois exemplos para chamar a atenção sobre o caráter essencial desses estudos. O primeiro exemplo envolve a noção de *proeminência* ou destaque de uma unidade em relação às demais, como vimos no caso da palavra "verde" em foco contrastivo.

O segundo exemplo envolve a noção de *segmentação* ou divisão da sequência falada numa série de agrupamentos. Tomemos a sequência de palavras, "Não", "comprou", "a" e "moto". A enunciação dessa sequência em dois agrupamentos sonoros, o primeiro formado por "não" e o segundo por "comprou a moto", significa que a pessoa anteriormente mencionada de fato comprou aquela moto. Por outro lado, a enunciação como um único agrupamento de palavras significa que a pessoa mencionada não comprou a moto. Essa distinção também é feita pela prosódia. Embora assinalar uma proeminência ou fazer uma segmentação não sejam as únicas funções prosódicas, os exemplos esclarecem ao leitor a importância do estudo desse componente da fala.

No que diz respeito aos estudos fonéticos, o pesquisador pode examinar tanto aspectos prosódicos na produção da fala quanto em sua percepção. Esses aspectos prosódicos podem ser observados porque são mensuráveis a partir de seus *correlatos* físicos. Correlatos são variáveis de um determinado nível que se relacionam ou variam conjuntamente com variáveis de outro nível. Assim, os correlatos físicos da prosódia são as medidas obtidas a partir do enunciado, chamadas de *parâmetros prosódicos* ou *prosódico-acústicos*, que assinalam a informação prosódica desse mesmo enunciado.

## 1.1 Correlatos físicos da prosódia

São três os principais correlatos físicos da prosódia: **frequência fundamental**, **duração** e **intensidade**. A esses três se pode acrescentar a **qualidade de voz**.

[1] As formas do substantivo "entonação" e do adjetivo "entonacional" também são possíveis e mais próximas do substantivo latino medieval *intonatio*. Optamos pelas formas mais vernáculas do português: "entoação" e "entoacional", acompanhando o paralelo entre "tom" e "entoar" (e não "entonar") e entre "som" e "soar".



Scanned with CamScanner

### 1.1.1 A frequência fundamental (F0)

A **frequência fundamental (F0)** é o equivalente acústico da frequência de vibração das pregas vocais (termo mais apropriado do que "cordas vocais", por razões anatomofisiológicas) e corresponde ao número de vezes em que as pregas oscilam em um segundo. Sua unidade física mais comum é o Hertz, abreviado Hz. A frequência fundamental também pode ser medida em **semitom** a partir de uma transformação logarítmica do seu valor em Hertz. Através desse artifício matemático, sua medida reflete melhor a forma como as vibrações são processadas por nosso sistema auditivo.

O semitom corresponde a um dos 12 intervalos da divisão da *oitava*, sendo que a oitava corresponde na música ao intervalo em que as frequências que o delimitam são o dobro uma da outra, Assim, o intervalo entre duas frequências na relação de oitava é de 12 semitons. O valor de uma frequência em Hertz pode ser convertido em semitom com referência à frequência de 100 Hz, cuja medida é abreviada "st ref 100 Hz", pelo uso de uma escala logarítmica como se vê na figura 1.1. A referência a determinada frequência é arbitrária. É uma decisão do pesquisador ou provém do uso consagrado em pesquisa. Significa decidir qual frequência será 0 semitom. Assim, quando 100 Hz é a referência, essa frequência será 0 semitom. Os valores em Hertz inferiores a 100 Hz terão valores negativos de semitom e os valores acima de 100 Hz terão valores positivos de semitons.

Assim, por exemplo, se o som é o lá fundamental, cuja frequência é 440 Hz, seu valor em semitons referidos a 100 Hz é de 25,65. Se usarmos a metade dessa frequência, 220 Hz, seu valor em semitons é de 13,65. Observe que a diferença entre esses números é de 12 semitons, exatamente uma oitava.

A importância do uso de uma escala logarítmica como o semitom está em duas características de nossa percepção auditiva. A primeira é o fato de nossa audição não perceber pequenas diferenças de frequência. O valor mínimo para a percepção, referido na literatura como *Just Noticeable Difference* (JND), isto é, a Diferença Minimamente Perceptível, é cerca de 1 semitom, que corresponde aproximadamente a 6 Hz. Mas esse valor é válido apenas para regiões até cerca de 1.000 Hz,

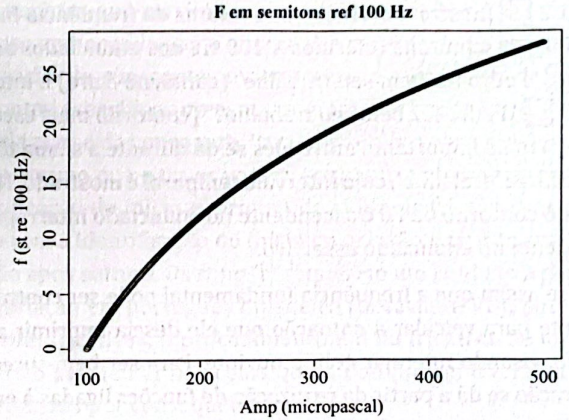


**Figura 1.1:** Relação entre frequência em Hertz (abscissa) e semitons ref 100 Hz (ordenada).
Equação 1.1 [12.log2(f/100)]

pois quando a diferença espectral, isto é, de frequência, se dá em faixas superiores a cerca de 2.000 Hz, nossa capacidade de perceber essas diferenças cai rapidamente. Essa segunda característica de nossa audição, a de discriminar melhor o som em baixa frequência e nem tanto em alta frequência, faz com que a expressão dessa discriminação de valores de frequência em semitom corresponda melhor a nossa percepção e é, assim, a medida mais apropriada para expressar frequência quando se trata do ser humano.

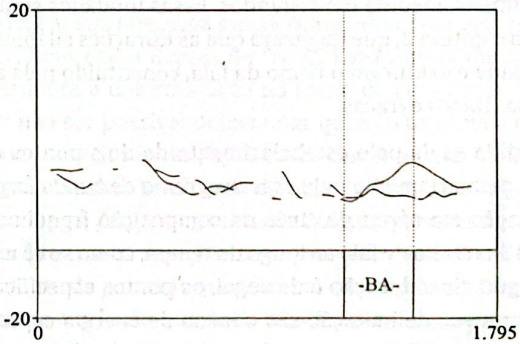


**Figura 1.2:** Curvas de F0 de enunciados assertivo "Pedro fez bem seu trabalho" (contorno claro) e interrogativo "Pedro fez bem seu trabalho?" (contorno mais escuro).
O eixo das ordenadas (vertical) é a F0 em semitons ref 100 Hz e o das abscissas, o tempo em segundos. O intervalo correspondente à tônica de "trabalho" está assinalado.



Scanned with CamScanner

A figura 1.2 mostra a evolução dos valores da frequência fundamental (F0) em semitons referidos a 100 Hz dos enunciados assertivo 1.4 "Pedro fez bem seu trabalho" (contorno claro) e interrogativo 1.5 "Pedro fez bem seu trabalho?" (contorno mais escuro). A diferença mais importante entre eles se dá durante a sílaba tônica "ba" da palavra "trabalho", cujo intervalo temporal é mostrado. Nesse intervalo, o contorno de F0 é ascendente no enunciado interrogativo e descendente no enunciado assertivo.

Observa-se assim que a frequência fundamental pode ser controlada pelo falante para veicular a entoação que ele deseja imprimir a seu enunciado, visando interagir com o ouvinte. Para ser bem-sucedida, essa interação se dá a partir da realização de funções ligadas à entoação da fala como no caso dos enunciados ilustrados aqui, bem como para assinalar acentos melódicos e segmentações da fala em unidades menores, entre outras funções, como será apresentado no capítulo 4.

### *1.1.2 A duração*

A **duração** se refere às unidades linguísticas que estruturam a informação prosódica dos enunciados. As medidas mais comuns são a duração de unidades silábicas, medida em milissegundos, e a duração de unidades superiores à sílaba, especialmente dos chamados grupos acentuais, medida em segundos. Essas unidades serão apresentadas no capítulo 2, que mostrará que as durações silábicas aliadas à intensidade estruturam o ritmo da fala, constituído pela alternância de sílabas fortes e fracas.

Sua medida se dá pelo estabelecimento de dois pontos específicos guiados primariamente pelo *espectrograma de banda larga*, uma representação em níveis de cinza da composição frequencial (ou espectral) de trecho de fala ao longo do tempo, como se vê na figura 1.3 a seguir. Se a duração é de vogal, os pontos específicos mencionados para sua delimitação são o início de energia espectral do segundo *formante* (F2) para marcar o início da vogal e o esmaecimento brusco da energia desse mesmo formante para marcar seu término. O formante é um modo de vibração obtido pela formação de ondas estacionárias que maximiza a energia de uma banda de frequência, por isso aparecendo como uma região escurecida no espectrograma. A duração de consoantes é determinada a partir da determinação de pontos específicos no espectrograma de banda larga que dependem da classe da consoante. Para as laterais e nasais, o procedimento é o mesmo que para as vogais, enquanto as fricativas requerem a identificação do início e término do ruído de fricção, e as oclusivas requerem a identificação do início de oclusão e término do ruído de fricção após soltura. Os sons "r" requerem um cuidado à parte, dada sua variação em português brasileiro (doravante PB), por exemplo. Se forem fricativos, o procedimento é o da fricativa. Se forem *taps*, é preciso segmentar as vogais que o flanqueiam e ter sua duração determinada por consequência.

A segmentação de vogais e consoantes é necessária também em pesquisa prosódica, uma vez que é preciso determinar segmentos acústicos correspondentes a sílabas, grupos acentuais, entre outros, que começam e terminam com palavras que envolvem diferentes vogais e consoantes.

Um exemplo de segmentação em sílabas e palavras pode ser visto na figura 1.3. Os critérios de segmentação de vogais e consoantes foram usados para determinar o início e fim das variadas sílabas desse trecho de "A Menina do Nariz Arrebitado", de Monteiro Lobato: "Em seguida apareceu um papagaio real que [...]", lido por um falante do interior de São Paulo com cerca de 20 anos no momento da gravação. É possível ver, em sua fala, uma pausa silenciosa entre "em seguida" e "apareceu", bem como depois de "real". Essa última pausa silenciosa impossibilita a determinação do início da consoante /k/ que a segue, por não ser possível determinar quando termina a pausa e quando começa a oclusão do /k/. Nesses casos, a determinação da duração dessa oclusiva, bem como da palavra "que", não é possível. À sílaba mais longa, /aw/, com cerca de 260 ms, precede uma pausa silenciosa, o que marca uma *fronteira prosódica*, isto é, o limite de um agrupamento de palavras que termina com a palavra "real". Esse exemplo sugere a relação que pode existir entre a presença de um parâmetro prosódico (a pausa), o valor de um parâmetro prosódico (duração alongada do ditongo e duração da pausa) e sua função prosódica (fronteira na cadeia de fala).



Scanned with CamScanner

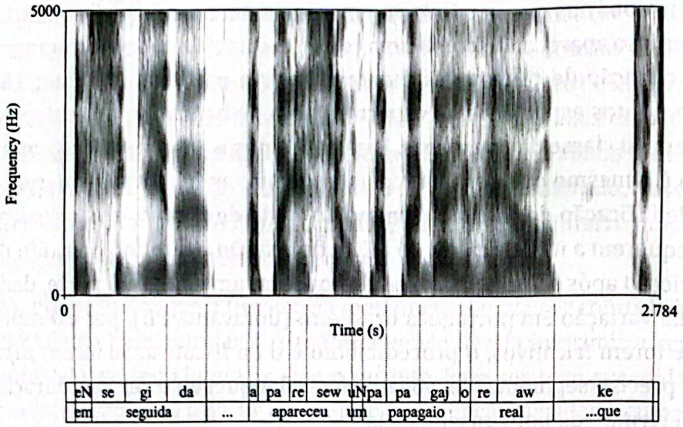


**Figura 1.3:** Exemplo de segmentação em sílabas e palavras (em baixo) no Praat a partir do espectrograma de banda larga (em cima). Observe no intervalo correspondente à sílaba /da/ a região escura mais inferior, onde atua o primeiro formante (F1) e a região escura logo acima, F2, referida no texto.

### 1.1.3 A intensidade

A **intensidade** de uma unidade prosódica pode ser **absoluta** ou **relativa**, mas, em ambos os casos, ela é proporcional ao quadrado da pressão sonora e expressa quão forte é um som. Calculada em toda a faixa de frequência de um som, a intensidade é dita absoluta. Se calculada a partir da diferença de intensidade em faixas de baixa e alta frequência, delimitadas por um limiar consagrado na literatura, a intensidade é dita relativa.

Por dupla conveniência, a intensidade é comumente expressa por uma unidade logarítmica, o **decibel**, abreviado dB. Procedendo assim, além de trabalharmos com números mais convenientes do que aqueles da medida de intensidade direta, nos aproximamos também da maneira como nosso sistema auditivo avalia a sensação derivada da variação de intensidade, o **volume**.

O decibel (dB) é definido em relação a uma amplitude ou intensidade sonoras absolutas de referência escolhidas em função do uso que se queira fazer da medida. Uma referência muito usada é a amplitude do *limiar de audição*, isto é, a amplitude mínima que alguém é capaz de ouvir. Uma representação gráfica da relação entre decibel e amplitude sonora pode ser vista na figura 1.4.

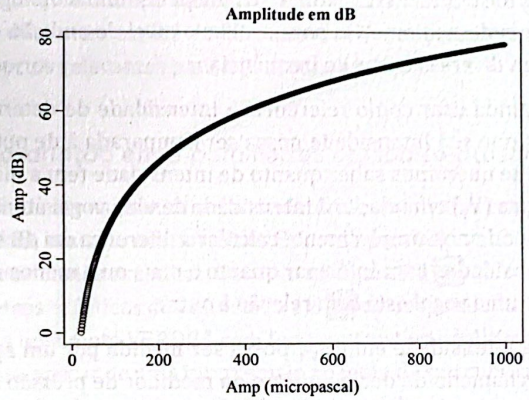


**Figura 1.4:** Relação entre amplitude em micropascal (abscissa) e decibel (ordenada) para uma amplitude de referência de 20 micropascais [Equação 1.2 Amplitude sonora em dB = $20.\log_{10}(I/I_r)$ I, é a intensidade sonora de referência e I é a intensidade cujo valor em dB se quer calcular]

Pela figura se pode ver que, até cerca de 200 micropascais, uma diferença em micropascais implica grandes diferenças em dB, mas, a partir daí, diferenças entre duas amplitudes em micropascais correspondem a diferenças bem menores em dB. Quando se usa o limiar de audição como referência, a unidade física é tecnicamente chamada de $dB_{SPL}$, em que SPL significa *Sound Pressure Level* ou nível de pressão sonora, isto é, quanto um som tem de pressão. Quando algum meio de comunicação se refere a, por exemplo, nível de intensidade de um avião de 130 dB, na verdade se trata de 130 $dB_{SPL}$.

A intensidade de referência pode também ser a de um limiar auditivo relativo a uma frequência específica, e não absoluta, como no caso do limiar de audição. A razão disso é que nossa sensibilidade ao som é distinta para frequências distintas. Em uma frequência de cerca de 1000 Hz somos mais sensíveis ao som. Em torno dessa frequência, situa-se o limiar de audição, mas, em frequências como 100 ou 10.000 Hz, somos pouco sensíveis e precisamos de som com intensidade bem mais alta do que a do limiar de audição para poder ouvi-lo nessas faixas. Por isso, podemos definir um tipo de cálculo



Scanned with CamScanner

em decibel com relação ao limiar de cada frequência e a essa medida chamamos de $dB_{SL}$, em que SL significa *Sensation Level*, isto é, o nível da sensação sonora. Essa medida é usada pelos fonoaudiólogos para fazer testes de audiometria para avaliar o nível de audição dos pacientes em diversas faixas de frequência.

Pode-se ainda usar como referência a intensidade de determinado som para que sua intensidade possa ser comparada à de outro. Por exemplo, se queremos saber quanto de intensidade tem a mais uma vogal tônica ($V_T$) em relação à intensidade de uma vogal átona contígua ($V_A$), podemos simplesmente calcular a diferença em dB entre as duas intensidades para informar quanto a mais ou a menos a intensidade de uma vogal está com relação à outra.

Embora a intensidade em $dB_{SPL}$ possa ser medida por um aparelho especial, chamado de decibelímetro ou medidor de pressão sonora, a medida da intensidade do sinal de fala por sua captação pelo microfone depende da distância desse com relação à boca. Assim, um cuidado experimental importante é a manutenção dessa distância o mais constante possível para fazer comparações entre medidas de intensidade que dependam apenas do uso que o falante faz desse parâmetro. Esse uso exige maior ou menor esforço vocal, que pode ser indiretamente medido pelo cálculo da intensidade relativa (ou ênfase espectral) em dB. Essa medida tem a vantagem de ser imune à posição do microfone com relação à boca. As medidas de intensidade relativa mais empregadas na literatura de prosódia da fala são as de *diferença espectral*, de *ênfase espectral*, entre outras, e envolvem a diferença de intensidade entre duas faixas de frequência, o que permite avaliar o esforço vocal do falante.

### 1.1.4 A qualidade de voz

A qualidade de voz (QV) está ligada à vibração das pregas vocais e à configuração laríngea durante um enunciado ou trecho de enunciado. Ela se manifesta acusticamente por modificações em aspectos como a regularidade e a tensão das pregas vocais, o que pode produzir uma voz crepitante ou a presença de ruído devido ao escape de ar, o que pode produzir uma qualidade soprosa, entre outras qualidades.

Essas diferentes qualidades, consequência de diferentes ajustes laríngeos, produzem efeitos no sinal de fala. O que justifica uma investigação da QV no campo de estudos prosódicos é o fato de ela poder marcar diferentes estilos de fala, atitudes e emoções do falante, todos aspectos relevantes para a prosódia da fala.

### 1.1.5 Covariação entre parâmetros prosódico-acústicos

Ao movimentar as estruturas articulatórias envolvidas com a produção da fala, mais de um correlato físico da prosódia varia conjuntamente em termos acústicos. Observe na figura1.5 duas curvas de parâmetros acústicos do enunciado 1.6 correspondente à sentença "Eu vi uma moto VERDE.", com foco contrastivo na última palavra. Vê-se que, apesar de uma interrupção no meio de seu curso por conta da realização da consoante /t/, durante a qual as pregas vocais não vibram, a curva de F0 (preta) começa e termina de crescer durante a palavra "moto". Após essa subida de F0, a curva cai bem rapidamente no início da palavra em foco, "verde", atingindo um valor baixo ao final da sílaba tônica "ver-". A curva de intensidade tem valores mais altos durante a palavra "verde" e na primeira sílaba de "moto". Além disso, a duração da sílaba focada, "ver-", é de 366 ms, a mais longa do trecho. O uso conjunto desses três parâmetros prosódico-acústicos guia o ouvinte para a percepção do foco na palavra "verde".

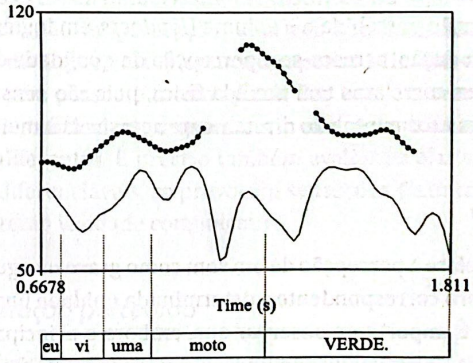


**Figura 1.5:** Curva de F0 (escura com pontos) e intensidade absoluta (linha contínua clara) do enunciado "Eu vi uma moto VERDE."



Scanned with CamScanner

Esse movimento de descida da F0 a um nível baixo durante a sílaba "ver-" permite assinalar ao ouvinte o caráter proeminente e contrastivo da palavra "verde", como se pode ouvir no exemplo dado. A participação de mais de um parâmetro prosódico-acústico para a implementação da função de proeminência contrastiva é comum. Assim, a realização conjunta, nesse mesmo exemplo, de intensidade absoluta com valores elevados reforça a relevância da palavra final. Se se guiasse apenas pela intensidade, o início da frase poderia ser considerado como igualmente importante, pois se vê que a intensidade também é elevada nesse momento. No entanto, a F0 é o parâmetro principal para o foco em trecho estreito, correspondente aqui a uma palavra, com a intensidade entrando como parâmetro de apoio.

Fica claro, então, que a forma como percebemos as variações dos parâmetros prosódicos e a forma como eles se conjugam é um mecanismo complexo que precisa ser examinado com mais vagar.

## 1.2 Correlatos perceptivos da prosódia

O falante controla os parâmetros prosódico-acústicos vistos acima a partir da articulação da fala para provocar no ouvinte efeitos perceptivos, isto é, uma sensação. Embora a cada um dos parâmetros vistos acima equivalha um efeito principal para a percepção, a relação não é de um para um. Os três correlatos perceptivos da prosódia são o *pitch*, a duração percebida e o volume (*loudness*, em inglês). A esses, como na produção, acresce-se a percepção da qualidade de voz. Nenhum desses correlatos tem medida física, pois são sensações, isto é, são efeitos no ouvinte não diretamente acessíveis à mensuração.

### 1.2.1 Pitch

O *pitch* se refere à percepção de um som como grave ou agudo em um trecho sonoro correspondente a determinada unidade linguística ou enunciado. É importante observar que, embora o principal parâmetro acústico que provoque a sensação de grave/agudo seja a frequência fundamental (F0), há dois cuidados importantes a esse respeito.

Em primeiro lugar, a sensação de *pitch* não cresce linearmente com essa frequência, mas de forma logarítmica. Daí o interesse em usar escalas como a de semitom, pois essa medida aproxima o valor em Hertz da sensação de *pitch*. Assim, à medida que a F0 aumenta, serão precisos valores cada vez maiores para sentir o mesmo grau de aumento de *pitch*.

Outro cuidado a tomar é: para mesmos valores de F0, mas de diferentes intensidades ou durações, a sensação de *pitch* é distinta, isto é, embora dependa primariamente da F0, o *pitch* é decorrente do concurso dos demais parâmetros prosódico-acústicos. A sensação de *pitch* de tons agudos aumenta com a intensidade, isto é, a sensação de agudo é tanto maior em tons agudos quanto maior a intensidade. Contrariamente, o *pitch* de tons graves decresce com a diminuição da intensidade. Outro aspecto composicional importante entre parâmetros acústicos e sensação é que, à medida que a duração diminui para o mesmo valor de F0, a sensação de grave aumenta, isto é, o *pitch* fica mais baixo. Por conta disso, é incorreto usar F0 e *pitch* indistintamente. O primeiro é da esfera da produção, portanto, mensurável, enquanto o segundo é da esfera de percepção, uma sensação sonora cujos efeitos se pode medir através da resposta de um sujeito.

Como vimos acima quando falamos de semitom, o ser humano também não é capaz de distinguir diferenças de qualquer valor entre duas frequências. Aquém de determinado limiar, essas duas frequências teriam o mesmo *pitch*. Frequências fundamentais com valores de 100 e 101 Hz são numericamente distintas, ou seja, são distintas na produção, não na percepção, uma vez que habitualmente uma pessoa não distingue uma diferença de apenas 1 Hz. Tais frequências terão para essa pessoa o mesmo *pitch*. Por isso, em termos experimentais, não basta mostrar que duas medidas prosódicas são objetivamente diferentes. É preciso também avaliar se elas são perceptivamente diferenciáveis, se provocam sensações distintas, de outra forma não terão validade comunicativa.

### 1.2.2 A duração percebida

A duração percebida é a sensação que permite avaliar se uma unidade é longa ou curta em relação a outra. O ser humano não é capaz,



Scanned with CamScanner

assim como para o *pitch*, de discriminar quaisquer disparidades entre valores de duração. Somos, por exemplo, mais sensíveis para apontar a diferença de duração entre duas sílabas átonas do que entre duas sílabas tônicas. Dito de outra forma, uma diferença de 40ms entre duas sílabas pode ser facilmente percebida se essas sílabas forem átonas, não percebida se forem tônicas. Se a comparação se dá entre sílaba tônica e sílaba átona, a percepção de maior duração da primeira é uma das propriedades acusticamente marcadas de sua tonicidade.

Essa sensação depende, portanto, da forma como os demais parâmetros acústicos variam ao longo de uma unidade linguística. É experimentalmente conhecido que ouvimos estímulos em que o valor de F0 varia como sendo mais longos do que aqueles em que F0 é constante, isto é, se o valor de F0 ao longo de uma vogal varia — sobe ou desce, por exemplo —, essa vogal é percebida como mais longa do que outra de mesma duração objetiva em que F0 não varia.

Da mesma forma que para a percepção de *pitch*, só se pode perguntar ao sujeito se uma unidade lhe parece mais longa ou curta do que outra ou se uma unidade parece durar o dobro (o triplo etc.) da outra. Mas não há como medir em unidade de tempo: é uma sensação. O que significa que pode variar entre os ouvintes também, devido a diferentes modos de perceber estímulos sonoros entre diferentes indivíduos.

### 1.2.3 O *volume* (loudness)

O volume é a sensação que se dá na escala forte/fraco. Ela é determinada primariamente pela intensidade de um som, mas não varia de forma linear com essa intensidade, mas, assim como para a frequência *vs. pitch*, à medida que a intensidade aumenta, serão necessários valores cada vez maiores para sentir o mesmo grau de aumento de volume. A relação entre intensidade e volume é, consequentemente, logarítmica.

Por conta do funcionamento de nossa orelha, a sensação de volume varia com a frequência do som e isso pode ser medido por testes com tons puros como mostra o resultado na figura 1.6.

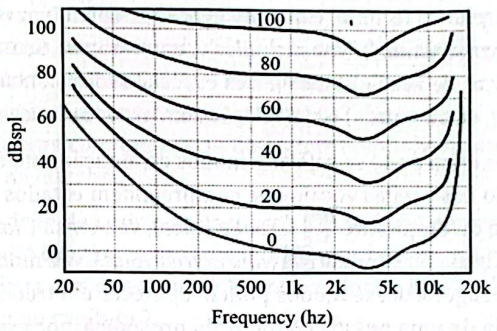


**Figura 1.6:** Contornos de mesmo volume segundo a frequência e intensidade ($dB_{SPL}$). Reproduzido de Zwicker e Fastl (1999).

Na figura acima, cada curva representa o mesmo volume, desde o limiar de audição percebido (curva mais baixa) até a mais alta, cujo volume está próximo do limiar da dor (100, em unidade arbitrária). A abscissa no gráfico representa a frequência do tom puro e a ordenada, a intensidade em $dB_{SPL}$. Observe então que, seguindo a curva vermelha, um tom de 50 Hz precisa de intensidade de 40 $dB_{SPL}$ para ser percebido como de mesmo volume que um tom de 1 kHz (1000 Hz) com intensidade de 0 $dB_{SPL}$. A figura mostra que o volume depende tanto da intensidade do som quanto de sua frequência, por conta da forma como nossa audição funciona. Da mesma forma, para um mesmo valor de intensidade, o volume aumenta com a alta da frequência.

A principal consequência dessa relação, no que tange à percepção da entoação, é o aumento de volume para uma mesma intensidade sonora à medida que o valor de F0 se eleva. Sendo assim, em trechos de ênfase com F0 crescente, o som tende a ser mais forte pelo fato de o volume aumentar também com a frequência.

### 1.2.4 A qualidade de voz percebida

A definição de qualidade de voz é matéria controversa, embora exista um consenso em propor tratar-se de uma resposta perceptiva a um sinal acústico não relacionada nem com acento de *pitch* nem com volume. É uma sensação auditiva relacionada a aspectos da voz compreen-



Scanned with CamScanner

didos com relativa facilidade quando evocados, como "voz rouca", "voz soprosa", "voz trêmula", "voz melodiosa", entre outros. Seu estudo envolve técnicas de avaliação por juízes especializados (fonoaudiólogos, foneticistas, entre outros) e sua relação com pistas acústicas.

O termo também pode se referir a modos de fonação, isto é, a modos de vibração das pregas vocais que compreendem estados de voz laringalizada ou crepitante (*creaky voice*), voz tensa (*harsh voice*), voz murmurada ou soprosa (*murmured voice*), voz modal, entre outros. Alguns desse modos podem aparecer em trechos específicos da fala de uma pessoa com função prosódica, por exemplo, demarcativa. A voz laringalizada, ao fim de um enunciado ou diante de uma pausa silenciosa, assinala, por exemplo, uma fronteira da cadeia de fala. Em outros casos, pode assinalar uma atitude em relação ao interlocutor, como o uso de uma voz tensa para marcar hostilidade.

## 1.3 *Softwares* para o estudo da prosódia

Na literatura, encontram-se tanto programas para análise do sinal de fala que podem ser usados para estudos prosódicos quanto programas específicos para esse tipo de estudo. O Praat e o WinPitch são exemplos de programas do primeiro tipo.

Desde seu lançamento em 1996, o Praat se constituiu como o *software* mais usado pela comunidade científica para análise acústica da fala, tanto segmental quanto prosódica. Acessível em <http://www.praat.org>, o programa integra uma série de funcionalidades que incluem aquelas que dizem respeito diretamente à análise prosódico-acústica como analisador de F0, analisador de intensidade e diversas medidas de qualidade de voz. Além disso, camadas de anotação podem ser associadas ao arquivo de áudio, permitindo segmentações de qualquer unidade linguística ou discursiva. A partir da combinação de anotação e áudio, podem-se obter medidas de duração, F0, intensidade e parâmetros relacionados à qualidade de voz para qualquer unidade prosódica de interesse. O programa permite ainda:

(1) montar testes de percepção de fala, que apresentaremos nas seções de procedimentos experimentais nos capítulos 3 e 4;
(2) filtrar o sinal de fala e manipular parte de seus parâmetros acústicos para estudar efeitos na percepção da fala;
(3) realizar algumas análises estatísticas;
(4) elaborar *scripts* para a automatização das análises prosódico-acústicas com uma série de *scripts* disponíveis em repositórios na internet.

A possibilidade de criar e utilizar *scripts* é um dos pontos fortes do programa e já existem muitos disponíveis para diversos tipos de análise. Uma lista dos que desenvolvemos ao longo dos anos pode ser encontrada no capítulo 7.

De interesse para os estudos prosódicos, o WinPitch, acessível em <http://www.winpitch.com>, também vem sendo desenvolvido desde 1996. Se, como o Praat, faz análise de F0 e possibilita anotação em camadas integradas ao arquivo de áudio, por outro lado, duas características que lhe são singulares são o fato de dispor de um *alinhador automático*, ferramenta que segmenta e etiqueta o sinal de fala em unidades do tamanho do fonema, e de permitir integração com arquivos de vídeo. Essa última funcionalidade constitui um ganho quando as ações são componentes importantes do estudo prosódico. Os pesquisadores da área cunharam o termo *prosódia audiovisual* para a investigação da relação entre aspectos prosódicos sonoros e visuais, veiculados essencialmente pela face. Semelhante às possibilidades de alteração de parâmetros acústicos do Praat pelo objeto Manipulation, o WinPitch permite o mesmo tipo de modificação, mas de forma gráfica, o que apresenta certa vantagem, uma vez que se podem fazer as modificações mais rapidamente quando não se dispõe de um *script* do Praat para tanto. É importante observar que o programa lê e salva arquivos de anotação em formato (TextGrid) usado pelo Praat. O programa vem sendo utilizado há muitos anos por equipes que investigam *corpora* de fala espontânea, como é o caso do projeto C-ORAL, de Cresti e Moneglia, e seus derivados, como o C-ORAL-Brasil, organizado há muitos anos por Raso e Mello.

Exemplos de programas específicos para estudos prosódicos são o Prosograma, de Mertens, e o Analor para o francês. Remetemos o leitor a informações sobre eles no *hotsite* dedicado a nosso livro no *site* da Parábola.



Scanned with CamScanner

CAPÍTULO 2

# Unidades prosódicas

A prosódia é o componente de nossa fala que organiza nossos enunciados, moldando nossa maneira de falar através do concurso de modificações articulatórias que se manifestam acusticamente em unidades prosódicas. Essas unidades organizam a nossa fala em níveis que vão da sílaba ao enunciado entoacional.

## 2.1 A sílaba

Em 1928, um grande pesquisador chamado Stetson sugeriu, a partir de investigação experimental da coordenação entre as atividades respiratória e de fala, que produzimos grupos respiratórios (*breath groups*) dentro dos quais a sílaba tem um papel constitutivo.

De fato, não é difícil perceber, quando alguém fala, uma oscilação mandibular com a qual podemos associar o abaixamento da mandíbula, mesmo de grau variável, à emissão da vogal e o seu alçamento à produção da consoante. Stetson chamava essa sequência de "trem de pulsos", em que cada pulso corresponderia a uma sílaba.

A sílaba é a unidade básica da fala, pois vogais e consoantes não ocorrem isoladamente. Apenas na sílaba. Assim, ao falarmos a sílaba /pa/ ou a sílaba /as/, consoante e vogal formam um todo, são produzidas numa única emissão de voz. Não produzimos o [p] para depois produzir o [a], mas os dois sons de forma integrada, num ciclo



Scanned with CamScanner

mandibular. O mesmo se pode dizer da sílaba /as/ ou de qualquer outra. Evidentemente, embora haja sílabas formadas por uma só vogal, como no caso do artigo "a", não há sílabas que sirvam para a comunicação linguística formadas por apenas uma consoante, como /p/, /f/, /m/ ou qualquer outra. Por vogais e consoantes fazerem parte de todas as línguas do mundo, ao constituírem sílabas, é natural cogitar a existência de uma sílaba "modelo", "típica", que denominamos de *sílaba canônica*. Essa sílaba canônica é a sílaba cv (consoante-vogal), como a sílaba /pa/ de nosso exemplo.

Há evidências experimentais para a *canonicidade* da sílaba cv, isto é, a preferência das línguas por essa estrutura, dentre as quais as mais importantes são:

- ✓ não há língua que não tenha sílabas cv e, para algumas, esta é a estrutura silábica mais frequente, como nas línguas românicas, com cerca de 50% de sílabas cv;
- ✓ sílabas cv são adquiridas em primeiro lugar durante a aquisição da linguagem. Pense em "papá", "dadá";
- ✓ sílabas cv são preservadas mesmo nas mais severas formas de afasia motora;
- ✓ processos diacrônicos de reestruturação silábica no curso de evolução de uma língua tendem a criar sílabas cv.

Outras evidências e experimentos que demonstram a canonicidade da sílaba cv podem ser encontradas no *hotsite* deste livro no portal de internet da editora.

A sílaba é uma unidade prosódica básica, isto é, a menor unidade de produção que somos capazes de emitir, podendo ser analisada do ponto de vista fonológico ou fonético. A *sílaba fonológica* pode ser estruturalmente definida como composta de um *ataque* e uma *rima*, como se vê na figura 2.1 com a sílaba representada por σ. O ataque é formado por uma ou mais consoantes, sendo simples, no primeiro caso, e complexo, no segundo. Assim, na sílaba /pa/, o ataque é simples e constituído pelo /p/, enquanto na sílaba /pra/, o ataque é complexo e formado pela associação de /p/ com o *tap*. Por sua vez, a sílaba /as/ não tem ataque, apenas a rima.

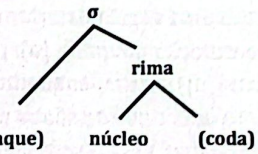


**Figura 2.1:** Estrutura silábica em ataque e rima. Os parênteses assinalam que o componente é opcional. Somente o núcleo é obrigatório.

O elemento mínimo da rima é o *núcleo*. Quando só composta por um núcleo, a rima é simples, enquanto, se associada a uma *coda*, a rima é ramificada. Nas sílabas /pa/ e /pra/ só existe o núcleo na rima, que é constituída pela vogal /a/. Já na sílaba /as/ ou na sílaba /mal/, a rima é ramificada e formada pelo núcleo constituído pela vogal /a/ seguida de uma coda: /s/, no primeiro caso, e /l/, no segundo.

Quando a sílaba é pronunciada, ela se realiza como *sílaba fonética* na cadeia da fala. É preciso ter em mente, no entanto, que os segmentos que constituem essa caracterização da sílaba nem sempre são realizados como sugere sua representação fonológica. Em seguida, vejamos dois exemplos de não correspondência direta entre sílabas fonológica e fonética.

A palavra "leite", formada por duas sílabas fonológicas, representadas normalmente pelas formas /'lej.te/ ou /'lej.ti/, pode constituir uma ou duas sílabas fonéticas. No caso de haver africação do /t/ e apagamento do /i/ final, que normalmente se realizaria como [ɪ], e admitirmos que a sílaba fonética pode terminar numa africada ([lejtʃ]), temos nesse exemplo uma única sílaba. Porém, se admitirmos que essa africada traz consigo um gesto palatal correspondente à realização de um [ɪ] ensurdecido, que seria o núcleo de uma segunda sílaba, temos então duas sílabas fonéticas ([lej.tʃ]). Em qualquer dos casos, não há, portanto, correspondência direta entre as representações silábicas fonológica e fonética, uma vez que, no caso apresentado aqui, há apagamento (*elisão*) da última vogal.

Na palavra "pacto", por sua vez, a representação fonológica dissilábica /'pak.to/ corresponde normalmente à realização fonética trissilábica ['pakɪtu]. Essa pronúncia indica que introduzimos uma vogal entre as duas consoantes mediais da palavra, o [ɪ], bem como reali-



Scanned with CamScanner

zamos o /o/ final como uma vogal posterior e alta pronunciada com menos intensidade e duração do que o [u], pois se trata do "u" que fazemos após a tônica ([ʊ]). Assim, enquanto no exemplo de "leite" a discrepância entre os dois tipos de sílaba provém da elisão da vogal [ɪ], no exemplo de "pacto" essa discrepância provém da inserção (*epêntese*) dessa mesma vogal. Para um aprofundamento da noção de sílaba fonética, especialmente no que se refere a sua segmentação [segmentação da silaba fonética], consultar o *hotsite* de nosso livro no portal de internet da editora.

Por não haver fronteiras claramente delimitáveis da sílaba fonológica na cadeia da fala, os foneticistas optaram por considerar outras unidades fonéticas que têm o tamanho da sílaba, mas não suas fronteiras. A unidade mais usada para esse fim é a *unidade V-V*, que começa no início acústico de uma vogal no núcleo silábico e vai até o início acústico da vogal no núcleo silábico imediatamente seguinte. Seria como "contar" as sílabas pelos seus núcleos, com a vantagem adicional de considerar o início da vogal, som com mais energia que os demais, como limites dessa unidade, o que a torna passível de segmentação automática. Supondo que o sintagma "as vacas de leite" seja pronunciado como [az vakɐz dʒi leitʃ], as unidades V-V seriam as seguintes, separadas por barras verticais: [az v|ak|ɐz dʒ|i l|ejtʃ]. Para mais detalhes sobre unidades VV , consultar o *hotsite* de nosso livro no portal de internet da editora.

Embora a sílaba seja uma unidade prosódica universal, isso não significa que outras unidades não possam ter um papel relevante na estruturação prosódica dos enunciados, como é o caso da **mora** em japonês, como se pode ver no *hotsite* de nosso livro no portal de internet da editora.

Outra evidência do caráter fundamental da sílaba como unidade de produção e percepção na fala é seu papel na versificação nas línguas românicas. Nessas línguas, a extensão dos versos e do texto poético se mede através da contagem do número de sílabas, um aspecto já presente nas formas orais da poesia como se encontra, por exemplo, entre os trovadores da baixa Idade Média no norte (*trouvères*) e sul da França (*troubadours*). A regularidade de extensão dos versos ajuda a memória, especialmente se se leva em conta que os versos eram cantados por esses artistas. E isso revela que, ao menos na poesia em língua românica, todas as sílabas contribuem para dar a noção de que os versos têm a mesma extensão. Já em línguas germânicas como o inglês, por outro lado, o verso se constrói pela contagem de pés, onde entra em jogo a relação entre *sílaba forte* e *sílaba fraca*. Uma sílaba é forte quando é percebida pelo ouvinte como mais importante do que as demais em sua vizinhança; ela é fraca se não é importante, havendo outras que se destacam. Em geral, em português, as sílabas tônicas são sílabas fortes, embora possa haver tanto sílabas tônicas que não são pronunciadas com força e, portanto, se comportam como fracas, quanto sílabas habitualmente átonas pronunciadas com força. No enunciado 2.1 , "Pedro se acostumou a brincar de bermuda clara.", as sílabas sublinhadas são fortes e as demais são fracas. Observe que a sílaba "-cos-" de "acostumou" é normalmente átona e aqui foi pronunciada com força (tornando-se um acento secundário), enquanto a sílaba tônica de "bermuda" foi pronunciada sem a força requerida para ser destacada no enunciado. O que faz uma sílaba ser forte ou fraca depende da forma de falar de cada um, numa determinada situação.

## 2.2 O pé

Se considerarmos as sílabas de palavras como "casa" e "casar", perceberemos imediatamente que, na primeira palavra, há uma relação de sílaba forte seguida de sílaba fraca, enquanto, na segunda palavra, a relação se inverte: sílaba fraca seguida de sílaba forte. O primeiro tipo de relação, forte/fraca, constitui um *troqueu* e, o segundo, fraca/forte, um *iambo*. Os dois são exemplos de pés. O **pé** é uma unidade que associa uma sílaba forte a uma ou mais sílabas fracas. Os mais comuns são os *pés binários* iambo e troqueu, que acabamos de ver. São também muito referidos na literatura, especialmente na versificação, os pés ternários *dátilo* (forte/fraca/fraca, como em "pérola") e *anapesto* (fraca/fraca/forte, como em "proteção"), que são os mais comuns, além do pé *anfíbraco* (fraca/forte/fraca, como em "careca").



Scanned with CamScanner

Embora a sequência de pés seja uma noção essencial para a criação de versos na versificação germânica, sendo frequentes na poesia inglesa tradicional os versos compostos por cinco iambos, a noção de pé para a prosódia da fala tem um papel menor em favor de unidades mais úteis para a organização dos enunciados, como a **palavra fonológica** e o **grupo acentual**. A principal razão disso é que, na enunciação, o número de sílabas átonas em sucessão pode ser muito grande, mesmo numa mesma palavra, como no caso de "paralelepípedo".

É fácil perceber que a sílaba mais forte dessa palavra é a antepenúltima, /pi/. Mas também não é difícil verificar que as quatro sílabas que a antecedem, em muitos casos, soam como pronunciadas por dois troqueus, (para) e (-lele-), seguidos de um dátilo (-pípedo). A maior força dessas sílabas átonas nos dois primeiros troqueus define o que chamamos de *acento secundário*, que assinala um grau menor de acentuação em relação ao grau com que a sílaba tônica é pronunciada, o local do *acento primário*. Outras pronúncias dessa palavra são possíveis, mas o acento primário sempre se mantém na posição da sílaba tônica da palavra. O que varia é o número de acentos secundários: um (paralelepípedo), dois (paralelepípedo) ou nenhum (paralelepípedo).

Esse jogo de acentos de diferentes forças com um único acento primário não se restringe aos limites do que comumente se chama de *palavra*, se estendendo para agregar palavras que não têm acento primário, os chamados *clíticos*. Clítico é uma palavra que precisa de outra, seu hospedeiro, para estabelecer uma relação de alternância acentual. É o caso de "a" na sequência "a casa", em que "casa", palavra lexical que tem acento primário em "ca", é seu hospedeiro. O clítico "a", formado por uma única sílaba, se comporta como uma sílaba pretônica do ponto de vista fonético. É fácil ver que ele parece formar uma única palavra com "casa" quando observamos pares de enunciados em que não há nenhuma diferença de pronúncia como em "Ele a bate" *vs.* "Ele abate". Pela pronúncia apenas não se podem distinguir os dois enunciados, que parecem ser formados por duas palavras. Daí a necessidade de definir uma noção que se assemelha à de palavra: a palavra fonológica.

## 2.3 A palavra fonológica ou prosódica

A palavra fonológica, ou *palavra prosódica*, é a unidade acima da sílaba que se pode considerar o máximo agrupamento que contém uma única sílaba portadora de acento primário e as sílabas átonas a ela associadas. Por exemplo, no enunciado 2.2 , "O rapaz canta para a moça.", as palavras funcionais "o", "para" e "a" são clíticos, definindo as palavras fonológicas "o rapaz", "canta" e "para a moça". Por serem, cada uma, uma palavra fonológica, elas têm apenas um acento primário, marcado pelo traço sublinhado. Como formam uma única palavra em termos prosódicos, elas podem estar sujeitas a processos de redução nas sílabas átonas. Assim, "para a moça" pode ser pronunciada "pra moça" pela elisão de vogais dos clíticos como ilustrado pela grafia "O rapaz canta pra moça."

Embora tenha definição fundamentada por critérios entoacionais, o chamado sintagma acentual (*accentual phrase*) proposto por pesquisadores ligados à teoria métrico-autossegmental, por se fundamentar na mesma restrição de um acento primário no constituinte, acaba sendo um sinônimo da palavra prosódica ou fonológica.

Acima dessa unidade, ainda por critérios de relação de força entre as sílabas, pode-se definir o grupo acentual, denominado em inglês de *stress group*.

## 2.4 O grupo acentual (*stress group*)

O grupo acentual está relacionado ao agrupamento de uma ou mais palavras fonológicas. Quando o agrupamento se dá com mais de uma palavra fonológica, apenas um dos acentos primários recebe papel de proeminência no enunciado. Por exemplo, no enunciado 2.3 , correspondente à sentença "A casa verde é bonita.", o primeiro sintagma é realizado com saliência acústica na palavra "verde", sendo "casa" desacentuada ou não saliente no enunciado. O trecho "a casa verde" constitui assim um grupo acentual formado por duas palavras fonológicas: "a casa" e "verde". Quando uma palavra fonológica recebe proeminência no enunciado de que faz parte, se diz que seu acen-



Scanned with CamScanner

to primário é frasalmente acentuado, isto é, é um **acento frasal**. É o caso do acento na sílaba tônica "ver-" da palavra "verde" do exemplo.

Portanto, o número de acentos frasais de um enunciado depende apenas da forma de pronúncia de uma sequência de palavras fonológicas. Assim, no caso do enunciado 2.2 🌐 "O rapaz canta para a moça.", pronunciado por dois agrupamentos em que somente a última palavra fonológica recebe acento ao nível do enunciado, "O rapaz canta" e "para a moça", há dois acentos frasais porque a palavra "rapaz" foi desacentuada para constituir o grupo acentual "o rapaz canta", em que apenas o acento primário de "canta" recebeu o acento frasal.

Por essa definição, o grupo acentual só pode ser determinado ao se ouvir o enunciado, pois a mesma sentença pode constituir um ou mais grupos acentuais, a depender da necessidade comunicativa do falante. Assim, a sentença acima pode ainda ser pronunciada pelos seguintes agrupamentos, que estão mais de acordo com uma divisão sintática entre sintagmas nominal e verbal: "O rapaz", "canta para a moça." Nesse caso, o primeiro grupo acentual, "o rapaz", tem acento frasal em "-paz" e o segundo grupo acentual, "canta para a moça", tem acento frasal em "mo-". Se a mesma sentença for emitida sem agrupamentos internos, teríamos um enunciado coincidindo com um grupo acentual de acento frasal na última sílaba com acento primário: "mo-".

Além da condição de ter uma única sílaba portadora de acento frasal, para a constituição do grupo acentual, é preciso saber como a língua interpreta o acento frasal como iniciando um grupo ou finalizando o grupo. Em línguas românicas como o português brasileiro, o francês e o italiano, os ouvintes têm a sensação de que o acento frasal termina o grupo acentual. Por isso, nos exemplos acima, definimos os grupos acentuais como finalizados pelo acento frasal, cuja sílaba portadora se encontra à direita das sílabas átonas. Em línguas germânicas, por outro lado, como o inglês, o alemão e o holandês, os ouvintes nativos têm a sensação de que o acento frasal começa o grupo acentual. E, portanto, esse acento se encontra à esquerda das sílabas átonas no grupo. Dizemos que as línguas em que o acento frasal termina o grupo são encabeçadas à direita (*right-headed*, em inglês) para essa unidade prosódica, enquanto dizemos que as línguas em que o acento frasal começa o grupo são encabeçadas à esquerda (*left-headed*, em inglês). O que contribui para essas duas diferentes percepções relaciona-se ao parâmetro acústico principal usado para marcar acento: a frequência fundamental nas línguas encabeçadas à esquerda e a duração nas línguas encabeçadas à direita.

Alguns autores definem grupo acentual (*stress group*) como sinônimo de palavra prosódica ou fonológica. Mas não vemos vantagem em tal escolha, pois ela não leva em conta a possibilidade de desacentuação das palavras fonológicas no enunciado.

Até esse nível, as unidades prosódicas envolvem noções de emissão de um único gesto de produção, no caso da sílaba, bem como de relações de força de emissão, para os demais casos: pé, palavra prosódica e grupo acentual. Essas relações são essencialmente rítmicas, como veremos no próximo capítulo. O uso do parâmetro da frequência fundamental para fins entoacionais nos permite definir unidades fundamentadas num contraste entre configurações dos contornos da curva de F0.

## 2.5 O sintagma entoacional (*intonational phrase*)

Um dos critérios para a delimitação de sintagmas entoacionais se relaciona à curva de F0, isto é, ao tipo de contorno de F0 que finaliza essa unidade e que pode ser ascendente ou descendente. A figura 2.2 🌐 mostra as curvas de F0 em semitons ref 100 Hz de dois enunciados distintos, o enunciado 2.2 🌐 "O rapaz canta/para a moça." e o enunciado 2.4 🌐 "O rapaz/canta para a moça." A barra inclinada (/) assinala a fronteira de dois sintagmas entoacionais em cada enunciado: no enunciado 2.2 eles são: "O rapaz canta" e "para a moça" e no enunciado 2.4, "o rapaz" e "canta para a moça". Observe na figura que o enunciado "O rapaz/canta para a moça." tem trecho da curva de F0 ascendente durante a sílaba final de "rapaz", assinalando uma fronteira não terminal, seguida de pausa silenciosa. Não é necessário, no entanto, como veremos no capítulo seguinte, haver pausa silenciosa para a percepção de uma fronteira.



Scanned with CamScanner

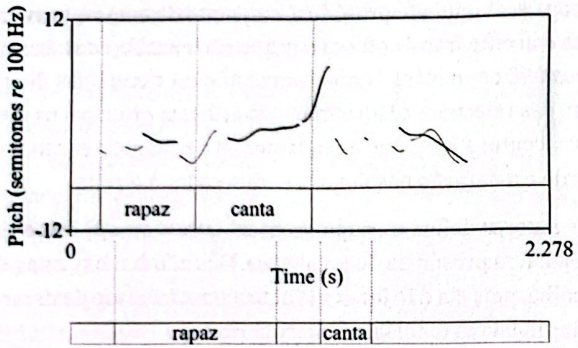


**Figura 2.2:** Curvas de F0 em semitons ref 100 Hz dos enunciados "O rapaz canta/para a moça." (linha clara) e "O rapaz/canta para a moça." (linha escura).

Já o enunciado "O rapaz canta/para a moça." tem trecho da curva de F0 ascendente durante a palavra "canta", assinalando uma fronteira não terminal também seguida de pausa silenciosa. Ambos os enunciados terminam praticamente juntos com um trecho de curva descendente que marca uma fronteira terminal, bem como o fim do segundo sintagma entoacional nos dois enunciados.

A forma que toma a curva de F0 permite delimitar os sintagmas entoacionais, possibilitando a percepção de agrupamento a partir dessas margens, assim consideradas pela sensação de fronteira terminal ou não terminal ligada à sensação de grave (curva de F0 descendente) e de agudo (curva de F0 ascendente). No entanto, parâmetros ligados ao ritmo da fala, como veremos no próximo capítulo, contribuem para a percepção de agrupamento e de unidade do sintagma entoacional, como a presença de pausa silenciosa e/ou alongamento de segmentos acústicos que antecedem a fronteira dessa unidade prosódica.

O sintagma entoacional é também referido na literatura de forma genérica como unidade entoacional (*intonational unit*) ou ainda como *sintagma fonológico* (*phonological phrase*) na literatura fonológica.

É importante ainda mencionar a possibilidade de usar critérios unicamente de produção da fala para associar os grupos acentuais a parâmetros de força, essencialmente a duração silábica e a intensidade, enquanto se podem associar os sintagmas entoacionais a parâmetros melódicos essencialmente ligados à forma ou chamados movimentos da curva de F0. Dito de outra forma, grupos acentuais seriam determinados pela saliência acústica de durações e intensidades de sílabas, enquanto sintagmas entoacionais o seriam apenas pelos movimentos de F0, fator primário para a organização entoacional da fala.

## 2.6 O enunciado (entoacional)

O enunciado pode ser definido por critérios pragmáticos como ligado a um ato de fala completo. Na literatura fonológica, é também denominado de enunciado entoacional (*intonational utterance*). Os enunciados 2.2 🌐 e 2.3 🌐 ilustrados acima realizam um ato de asserção que se termina com uma fronteira, no caso, a cessação da fala, denominada de fronteira terminal.

A mesma sequência de palavras pode constituir enunciados distintos cuja função comunicativa só pode ser recuperada pelo ouvinte a partir da prosódia. Assim, a sequência de palavras "João", "faz, "isso", pode formar:

(1) dois enunciados, sendo um vocativo seguido de um enunciado de ordem no exemplo 2.5 🌐;
(2) um enunciado único, uma asserção, como no exemplo 2.6 🌐;
(3) um enunciado único, uma questão, como no exemplo 2.7 🌐

que podem ser ouvidos no link deste livro no *site* da Parábola.



Scanned with CamScanner

CAPÍTULO 3

# O ritmo da fala

O ritmo da fala pode ser definido pela conjugação de duas percepções: a de regularidade silábica e a de alternância de sílabas fortes e fracas. Já em 1919, Whitehead dizia: a essência do ritmo é a fusão da identidade e da novidade. Essa identidade se refere à mencionada regularidade, enquanto a novidade se refere às sílabas fortes, que chamam a atenção do ouvinte ao emergir de um "fundo" de sílabas fracas.

Esse concurso de mecanismos contrários, um tendente a igualar sílabas e outro tendendo a diferenciá-las, é comum nas canções e nos poemas e contribui para a fixação da arte musical e literária. Nos versos a seguir, de Guilherme de Almeida, se pode ver o jogo entre sílabas átonas, declamadas de forma regular, e as sílabas fortes, que se destacam para o ouvinte e são aqui assinaladas pelo traço sublinhado: "Pela primeira vez eu comecei a ver, / dentro da própria vida, o encanto de viver."

As sílabas fortes marcam os acentos frasais que, em nossa língua, terminam os grupos acentuais. No exemplo acima, os grupos acentuais seriam os delimitados pela barras verticais (com duas barras, marcamos a fronteira entre os versos) e logo abaixo o número de sílabas em cada grupo:

| Pela primei \| ra vez \| eu co \| mecei a ver, \| \| dentro da pró \| pria vi \| da, o encan \| to de viver. \| |
|---|
| 4 2 2 4 4 2 4 4 |

Vê-se claramente uma combinação de grupos acentuais de 4 e de 2 sílabas que provoca um efeito de facilidade de declamação que contri-



Scanned with CamScanner

bui para a memorização dos 8 grupos acentuais dos dois versos. Observe que as sílabas fortes ocorreram em todas as sílabas tônicas das *palavras lexicais*, com exceção do deslocamento da força de emissão da tônica para a sílaba inicial de "comecei", que realiza então um acento secundário. O trecho de sílabas até a primeira sílaba forte de cada verso é referido como *anacruse*. Mais detalhes sobre isso no *site*.

## 3.1 Línguas de ritmo silábico e de ritmo acentual

Na década de 1940, o foneticista galês Arthur Lloyd-James, estudando o sinal de fala na transmissão telefônica, se referiu aos ritmos percebidos do espanhol e do inglês como respectivamente "ritmo de metralhadora" e "ritmo de código Morse". Ele fazia alusão à metralhadora, no primeiro caso, provavelmente por conta da incapacidade de perceber uma distinção entre sílabas fracas e fortes, por ser uma língua que lhe era estrangeira. O ritmo de código Morse foi atribuído ao inglês pelo contraste percebido entre sílabas longas (as fortes) e curtas (as fracas). Essa observação conduziu à percepção de que as línguas teriam ritmos distintos que se agrupariam pela forma como as sílabas eram percebidas, cunhando-se os termos **ritmo silábico**, se as sílabas parecessem ter todas a mesma regularidade, e **ritmo acentual**, se as sílabas fortes fossem percebidas como regulares, independentemente do número de sílabas fracas entre esses momentos fortes.

Mesmo que na declamação de versos em línguas germânicas os pés tendam a ser regulares e, assim, a favorecer a percepção da regularidade de sílabas fortes e mesmo que a declamação de versos em línguas românicas tenda a valorizar cada sílaba, é fato que a regularidade, tanto de uma sequência silábica sem distinção de força, quanto de uma sequência de sílabas fortes, não se dá na fala. O principal fator para essa não regularidade absoluta é que a realização de acentos frasais nos enunciados exige a diferenciação dos parâmetros acústicos, notadamente duração e intensidade, relacionados à percepção de força, o que quebra a estrita regularidade. De outra forma, o acento não seria perceptível para o ouvinte.

Ainda não se sabe ao certo onde residiria a sensação desses dois tipos de ritmo, aos quais, para algumas tipologias, acresce-se o *ritmo moraico* de línguas como o japonês, que se fundamentam na mora como unidade básica da constituição do ritmo da fala. A mora é uma unidade menor que a sílaba, uma forma de mensurar sua extensão. Para mais detalhes, consultar o *site*. Pode ser que essa sensação de ritmos distintos resida na conjunção de uma série de fatores que podem estar relacionados ao grau de redução vocálica maior em línguas de ritmo acentual, que pode contribuir para uma discrepância maior entre sílabas fortes e fracas ou ainda conjuntamente pela importância do acento de *pitch* que, por envolver essencialmente o uso da F0, exacerba a maior saliência acústica da sílaba forte, contribuindo para uma possível regularidade da ocorrência de sílabas fortes ao longo da cadeia de fala. Na seção 3.4, voltaremos ao tema da regularidade silábica apontando aspectos de diferenciação na duração de unidades silábicas. Para saber mais sobre tipologia rítmica, consultar o *site*.

## 3.2 Funções rítmicas de segmentação e proeminência

Por serem duas funções prosódicas fundamentais, a segmentação e a proeminência são assinaladas através do ritmo da fala a partir da duração de unidade do tamanho da sílaba. Do ponto de vista funcional, isto é, com papel na comunicação falada, vimos na introdução que a *segmentação prosódica* diz respeito à capacidade de reconhecer unidades constitutivas da fala. No que tange ao ritmo da fala, as unidades mais relevantes estão associadas à percepção de fronteiras terminais e não terminais na cadeia de fala. Essas fronteiras estão associadas à percepção de um acento frasal que assinala respectivamente terminalidade ou não terminalidade, a depender de um conjunto de parâmetros acústicos que concorrem para o assinalamento de acento frasal, tendo a duração o papel mais relevante. As unidades constitutivas, ou *constituintes prosódicos* associados à percepção de fronteira ligada à duração, são os grupos acentuais. Como veremos na seção 3.5, do ponto de vista perceptivo, a delimitação experimental de constituintes prosódicos é feita a partir de um conjunto de ouvintes.



Scanned with CamScanner

A função de proeminência é a sensação associada ao destaque de uma unidade linguística em relação a outras de mesmo tamanho em sua vizinhança imediata. Essa proeminência, em português brasileiro, também se relaciona de maneira mais determinante à percepção local de duração mais longa da sílaba de uma palavra, normalmente sílaba sentida como forte e que realiza o acento primário ou um ou mais acentos secundários da palavra. Em testes de percepção, podemos perguntar ao ouvinte quais palavras ou sílabas se destacam num trecho de fala, sendo a palavra a unidade mais eficiente.

A função de proeminência está relacionada a dois outros termos, os de foco e ênfase, sendo constante a confusão entre os três.

### 3.2.1 Foco e ênfase

O **foco** diz respeito à estrutura informacional de uma sentença, sendo uma unidade linguística em destaque por conta de uma série de motivos tais como: ser informação nova, elemento recorrente numa narrativa, elemento em contraste na mente dos interlocutores, como no caso do foco contrastivo [visto na Introdução]. O foco é assim definido tanto por teorias de natureza sintática quanto semântica, tendo domínios de diferentes extensões. A esse respeito, e com relação à interface com a prosódia, convém falar de *foco estreito* e *foco largo*.

Do ponto de vista prosódico, o foco estreito é sempre realizado de forma a suscitar uma proeminência que assinala a unidade em foco. Assim, no enunciado 3.1 🌐, aqui reproduzido: "Eu vi uma moto VERDE.", a palavra "verde" está em foco e, por a palavra ser uma unidade menor que o sintagma, o foco se diz estreito. O foco estreito pode se dar até numa sílaba, como no enunciado 3.2 🌐, correspondente à sentença "Não, Maria está INfeliz.", com foco estreito na sílaba "in-", como resposta a uma questão como "Maria está feliz?".

No caso do foco largo, mesmo estando em foco como unidade, sintática ou semanticamente, uma sentença inteira ou um sintagma, a proeminência se realiza numa unidade menor que a do trecho em foco, numa posição padrão de realização de acento frasal, qual seja, a última palavra fonológica à direita do trecho, no caso do português brasileiro. Assim, no enunciado 3.3 🌐, correspondente à sentença "João comprou [uma casa verde]$_F$.", como resposta a uma pergunta sobre o que João comprou, o sintagma em foco largo, aqui marcado pelo subscrito "$_F$" no trecho entre colchetes, se realiza com proeminência na palavra "verde" apenas, por ser esta a palavra mais à direita no sintagma.

O termo **ênfase** diz respeito a uma proeminência associada a um foco estreito. É a realização, na fala, de um foco estreito definido nos domínios sintático ou semântico. Se a ênfase realiza o foco contrastivo, ela é chamada de *ênfase por contraste*, remetendo a um elemento ausente na cadeia da fala. Se, por outro lado, destaca uma unidade linguística no eixo sintagmático chamando sua atenção em relação ao que é dito no próprio enunciado, se diz *ênfase por intensidade*.

Para além dos aspectos ligados a proeminência, segmentação e tipologia do ritmo, um importante aspecto da fala está ligado a quão rápida ou lentamente determinado falante produz seus enunciados, conforme segue.

## 3.3 Taxa de elocução (*speech rate*) e de articulação (*articulation rate*)

A taxa de elocução, *speech rate* em inglês, é o número de unidades linguísticas por unidade de tempo produzidas por um falante. Suas medidas mais comuns são "palavras por minuto" e "sílabas por segundo". Como a primeira pode dar uma ideia falsa da distinção entre taxas com que dois falantes de diferentes línguas enunciam — pelo fato de as línguas terem extensão de palavras distintas —, recomenda-se a mensuração em sílabas por segundo.

A denominação "taxa de elocução" no lugar do uso vulgar de "velocidade de fala" é preferível, por apontar para o que de fato ocorre: a produção de um número de unidades por intervalo de tempo, evitando a ideia errônea de que os articuladores da fala estariam se movendo mais rapidamente. De fato, ao acelerarmos nossa fala, fazemos excursões menores com os articuladores, não produzimos determi-



Scanned with CamScanner

nados segmentos, fundimos dois segmentos num único, ajustes que permitem fazer caber mais unidades num mesmo intervalo de tempo.

A taxa média dos falantes é a produção de 4 a 6 sílabas por segundo, mas esse número pode variar com relação a aspectos como:

(1) estilo de elocução (leitura, narração, locução profissional);
(2) gênero da produção (aula, conversa, entrevista, discurso);
(3) grau de familiaridade entre interlocutores;
(4) grau de formalidade da produção;
(5) idade;
(6) escolaridade;
(7) gênero do falante, entre outros.

Tendo em vista que a sílaba pode ser definida tanto ao nível fonético quanto ao fonológico, o pesquisador deverá informar de que forma as sílabas foram consideradas: sílabas fonológicas por segundo ou sílabas fonéticas por segundo. Para o cálculo da taxa de elocução basta, então, contar esse número e dividi-lo pelo tempo total (em segundos) em que elas foram enunciadas, incluindo as pausas silenciosas produzidas pelo falante. Essa distinção tem implicações importantes. Para uma mesma sentença pronunciada por diferentes falantes, o número de sílabas fonológicas é sempre o mesmo, mas o de sílabas fonéticas depende da realização silábica de cada falante, pois um falante pode elidir sílabas e outro não, um falante pode produzir um elemento epentético e outro não. Assim, mesmo que suas taxas de articulação em sílabas fonológicas fossem as mesmas (por falarem no mesmo tempo) entre os falantes, a depender da presença de um fenômeno de elisão ou epêntese num deles e no outro não, as taxas de articulação em sílabas fonéticas seria distinta. É preciso fazer a escolha por um tipo de sílaba por alguma razão teórica e ser consistente, medindo sempre da mesma forma todos os dados.

Se quisermos avaliar a taxa de produção apenas quando há fonação, isto é, excluindo os trechos de silêncio relacionados a pausas, usamos uma medida que se denomina **taxa de articulação** (*articulation rate*), calculada pela divisão do número de sílabas (fonológicas ou fonéticas) pelo tempo de enunciação, excluindo-se dele a soma das durações das pausas silenciosas produzidas no trecho de medida da taxa. Se nenhuma pausa for produzida, as duas medidas de taxa terão o mesmo valor.

Em seguida, passaremos a apresentar procedimentos experimentais para a pesquisa sobre ritmo da fala, apontando os cuidados metodológicos essenciais.

## 3.4 Procedimentos experimentais da pesquisa sobre ritmo da fala

A pesquisa sobre o ritmo da fala concerne essencialmente à observação, à descrição e ao modelamento das durações silábicas dos enunciados de uma língua para o estudo das funções prosódicas como acento frasal, proeminência e segmentação, bem como para o estudo das consequências sobre o ritmo da fala de mudanças de estilo de elocução (leitura, narração, conversa, fala de jornalista, entre outros), de gênero (narração para crianças, narração esportiva, comentário jornalístico ou manchete, entre outros), de nível de formalidade entre interlocutores, de mudanças no estado emocional, de atitude, entre muitas outras situações. Exemplos com dados de pesquisa real são usados nesta seção para a orientação do leitor.

### *3.4.1 A duração silábica e a estruturação rítmica*

Para mais precisamente relacionar a medida objetiva da duração silábica com sua função rítmica, recomenda-se partir da duração da unidade V-V, que considera o início da vogal e não da sílaba fonológica como fronteira de unidade silábica para fins de estudo prosódico. Por exemplo, na sequência /kaza veRde/, as sílabas fonológicas são /ka/, /za/, /verR/ e /de/, enquanto as unidades V-V seriam apenas as que se podem delimitar claramente entre inícios de vogal: /az/, /av/ e /eRd/. No *site* se encontram resultados de pesquisas psicolinguísticas que justificam seu uso no lugar da sílaba fonológica.

Mas, em unidades de tempo, não basta o uso da duração bruta, pois a duração percebida se dá quando seu valor difere de um valor "esperado", correspondente aos valores de duração médios que ouvimos



Scanned with CamScanner

das unidades silábicas. Portanto, para revelar aspectos funcionais, é preciso normalizar a duração bruta.

Para ilustrar a obtenção de medidas de duração normalizada da unidade V-V a partir de três etapas, comecemos mostrando o contorno de duração normalizada e suavizada das unidades V-V do enunciado 3.4 "Manuel tinha entrado para o mosteiro há quase um ano, mas ainda não se adaptara àquela maneira de viver." na figura 3.1., realizado por falante paulista do sexo feminino de cerca de 30 anos na época da gravação.

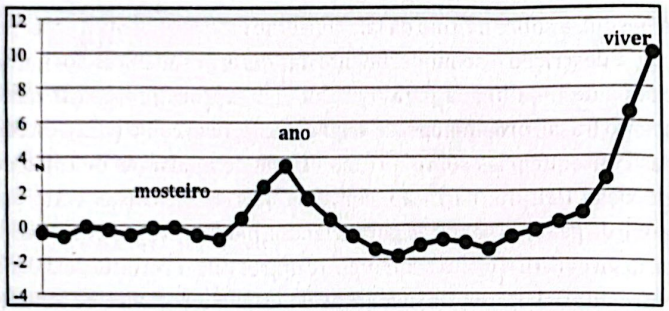


**Figura 3.1:** Contorno duracional das unidades V-V do enunciado "Manuel tinha entrado para o mosteiro há quase um ano, mas ainda não se adaptara àquela maneira de viver." por uma informante paulista.

Para a obtenção do contorno da figura 3.1, usou-se um procedimento em três etapas:

(1) delimitação e etiquetagem das unidades V-V;
(2) normalização das durações com o método de z-*score*;
(3) suavização dos z-*scores* por método de média móvel de 5 pontos.

Esse procedimento em três etapas suprime a medida física temporal em milissegundos pela obtenção do valor de "z" da ordenada que é uma medida relativa que expressa em quantas unidades de desvio-padrão a duração da unidade V-V está afastada de uma média de referência. A etapa de suavização permite atenuar picos locais de duração sem relevância prosódica. O procedimento permite revelar, com acuidade de cerca de 80%, aquilo que é percebido como saliente pelos ouvintes, seja pela marcação de uma fronteira ou pelo assinalamento de uma proeminência. Observe na figura que os dois principais picos de duração na figura se encontram nas tônicas das palavras "ano" e "viver" que assinalam o fim das orações coordenadas. A tônica da palavra "mosteiro" também tem duração com pico de menor altura, que contribui para a percepção de que essa palavra é proeminente. Há ainda outros dois picos menores (em "entrado" e em "adaptara") que não são percebidos por todos os ouvintes e, por isso, não são assinalados por suas palavras na figura.

A primeira etapa é ilustrada pela figura 3.2, que mostra o espectrograma de banda larga, a segmentação em unidades V-V e a etiquetagem do trecho "(most)eiro há quase um ano, m(as ainda)" do enunciado 3.4 "Manuel tinha entrado para o mosteiro há quase um ano, mas ainda não se adaptara àquela maneira de viver."

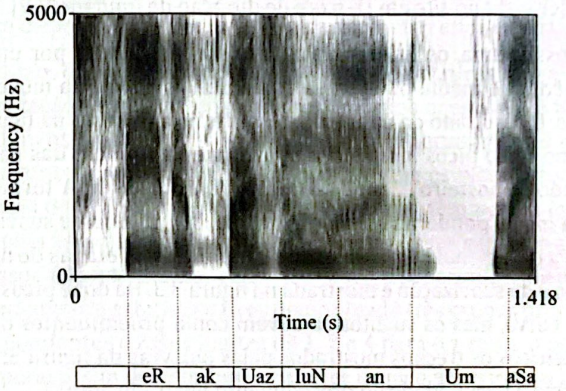


**Figura 3.2.** Espectrograma de banda larga acima, segmentação em unidades V-V e etiquetagem do trecho "(most)eiro há quase um ano, m(as ainda)" do enunciado exemplo referente à figura 3.1.

A delimitação das unidades V-V segue a identificação dos inícios das vogais, que assinalamos em negrito pela transcrição fonética com o AFI[1]: [**e**r **a** k**wa**z**ɪ̯ũa**n**ʊ** mas]. Juntando as semivogais subjacentes ([w]) ou tornadas tais para formar um ditongo a partir da sílaba final de "quase" e o artigo "um", têm-se as seguintes unidades V-V: |**e**r |**a** k|**wa**z|**ɪ̯ũ**|an|**ʊ** m|. A vogal final do trecho (/a/ em "aSa", de "mas a(inda)") inicia a unidade V-V seguinte, que termina na próxima vogal

[1] Os símbolos fonéticos na figura 3.2 são etiquetados com caracteres ASCII por conta de uso computacional.



Scanned with CamScanner

não mostrada no espectrograma, por isso apenas 6 unidades são definidas no trecho. São as durações dessas unidades e os símbolos fonéticos de cada uma que são usad0s como entrada para a etapa seguinte.

Na etapa de normalização, calcula-se um valor normalizado de duração que é padrão na estatística, o *z-score*. Por definição, esse valor expressa o grau de afastamento do valor bruto de duração em relação à média em unidades de desvio-padrão. Dizer, por exemplo, que a duração de uma unidade ou sílaba tem valor de z-*score* igual a 1,3 quer dizer que a duração se encontra a 1,3 desvio-padrão acima da média da duração. Os valores de referência de média e desvio-padrão de duração podem ser obtidos de um *corpus* que não precisa ser com a fala do mesmo informante[2]. A fórmula empregada para essa etapa pode ser vista no *site* em [z-*score* de duração de unidade V-V].

Após essa etapa, os valores de z-*score* são suavizados por um procedimento de média ponderada conhecida como média móvel de 5 pontos. O resultado da aplicação dessa técnica foi visto na figura 3.1 e define cinco picos locais correspondentes às tônicas das palavras "entrado", "mosteiro", "ano", "adaptara" e "viver". A fórmula da média móvel ponderada se encontra no *site* em [z-*score* suavizado]. A curva duracional com a duração bruta, antes das etapas de normalização e de suavização é mostrada na figura 3.3. Há doze picos locais nessa curva, mas os sujeitos só ouvem como proeminentes ou pré-fronteiriços os trechos mostrados pelas palavras da figura 3.1. Isso confirma a necessidade do procedimento ilustrado aqui.

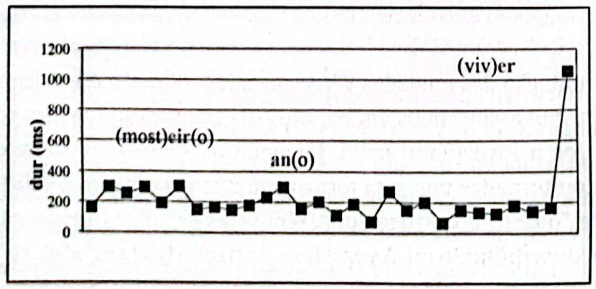


**Figura 3.3:** Contorno duracional bruto referente ao enunciado da figura 3.1.

[2] Cf. Barbosa, 2006: 489, para uma tabela de referência.

Os picos de duração ao longo do enunciado ilustrado nesta seção, mostrados na figura 3.1, podem ser usados como critério para a delimitação de grupos acentuais, como definidos na seção 2.4. Sendo assim, no enunciado 3.1, os grupos acentuais delimitados por esse critério de duração seriam os seguintes: [Manuel tinha entr**ad**] [o para o most**eir**] [o há quase um **an**] [o mas ainda não se adapt**ar**] [a àquela maneira de viv**er**.] Embora deva-se privilegiar sua delimitação por critérios perceptivos, como veremos na próxima seção, tendo em vista a correlação em geral superior a 80% dos picos de duração normalizada com a percepção de fronteira ou proeminência na palavra em que incidem esses picos, podem-se considerar esses picos limites à direita de grupos acentuais. A vantagem desse critério é tornar o processo de obtenção de grupos acentuais automático, desde que se tenham disponíveis as unidades V-V devidamente etiquetadas.

Além disso, os valores distintos em altura desses picos possibilitam a delimitação de unidades superiores ao grupo acentual a partir de procedimentos estatísticos que apontem para agrupamentos numéricos distintos. Pode-se ver, pela figura 3.1, que os picos associados às tônicas de "entrado", "mosteiro" e "adaptara" têm z-*score* inferior a 1, o pico associado à tônica de "ano" tem z-*score* superior a 2 e o pico associado à tônica de "viver" tem z-*score* superior a 4. Se o teste de agrupamento estatístico tiver determinado três grupos principais correspondentes a esses limites de 1, 2 e 4 para o z-*score*, o pesquisador pode assim decidir: que os valores de pico superiores a 2 e inferiores a 4 coincidiriam com fronteiras de sintagmas entoacionais, bastando para isso verificar se um tom de fronteira melódica está associado a esse limite (veremos no capítulo seguinte ser exatamente o caso aqui) e que os valores de pico superiores a 4 coincidiriam com fronteiras de enunciado, que é o caso aqui também. Esses números-limite são apenas exemplos que necessitam de verificação experimental em *corpora* amplos. Observe-se que técnicas estatísticas que permitem realizar esse tipo de separação de conjuntos de valores requerem grande quantidade de dados para apresentarem resultados relevantes e que, portanto, essas técnicas devem ser empregadas para um conjunto considerável de enunciados de um mesmo falante. As unidades superiores de sintagma entoacional e enunciado, como



Scanned with CamScanner

vimos anteriormente, são constituídas por grupos acentuais com diferentes graus de força da fronteira que podem ser representadas pelos valores dos picos de duração normalizada.

Os picos de duração locais podem também ser usados como caracterizadores de um estilo de elocução ou de um ritmo da fala num mesmo falante ou em falantes distintos. Se definirmos uma distância entre conjuntos de valores pela diferença entre médias desses valores em unidades de desvio-padrão médio podemos revelar diferenças rítmicas entre trechos de fala. Ouça no *site* os pares de enunciados por falantes distintos do trecho 3.5 "O mesmo não aconteceu, porém, com a poetisa Lagartixa, que principiou a recitar uma longa poesia e engasgou no meio, acabando o recitativo em choro e faniquito." A distância entre valores de z-*score* na leitura dos dois falantes que soam falar mais distintamente é de 3,48 enquanto a daquela em que os falantes soam mais semelhantes é de 1,14. Tem-se aí um grande potencial para apontar diferenças rítmicas sem a necessidade de perguntar sistematicamente aos ouvintes o que demanda muito tempo e investimento de pesquisa. No entanto, a investigação de funções prosódicas básicas associadas ao ritmo por meio de testes de percepção é um componente importante da pesquisa experimental sobre ritmo da fala.

### 3.4.2 Elementos para estudo experimental da percepção da prosódia da fala

A determinação de funções de natureza prosódica, como a marcação de proeminências e de fronteiras na fala, pode ser feita pela escuta de trechos de fala de uma bateria de ouvintes, aos quais se solicita marcar as palavras que cumpram uma das duas funções básicas, instruindo-os a considerar se são proeminentes ou se precedem imediatamente uma fronteira. Alguns autores solicitam ao ouvinte tarefas mais árduas, como apontar sílabas proeminentes ou indicar graus de proeminência de palavras ou sílabas, mas os resultados desse tipo de refinamento exigido do ouvinte são muito variáveis, logo, não recomendados. A vantagem de uma tarefa binária é evidente e o grau de proeminência ou força da fronteira pode ser obtido a partir da porcentagem de respostas dos ouvintes para palavras pré-fronteiriças ou proeminentes. Vejamos um exemplo.

O trecho 3.6 , lido por um jovem universitário paulista, foi extraído do livro *A menina do nariz arrebitado*, de Monteiro Lobato. Sua transcrição ortográfica mostra, entre parênteses após uma determinada palavra, a porcentagem de 10 ouvintes que a perceberam como proeminente (quando nenhuma porcentagem é assinalada, nenhum dos ouvintes a considerou proeminente).

> Em (20) **seguida** (60) apareceu (10) um papagaio (20) **real** (90) que tinha fama (20) de **orador** (60). Subiu (10) a tribuna (20) de um poleiro (20) de **ouro** (50) e fez um **belo** (90) discurso (20) a respeito (20) da **arte** (60) de **falar** (30). Nesse (10) **discurso** (30) provou (10) que os homens (20) **tinham** (40) aprendido a falar (10) com os **papagaios** (50), e **não** (50) os papagaios (20) com os homens (20), como diz (20) a **ciência** (30) **destes** (40). Uma chuva (10) de **palmas** (50) acolheu (10) suas **palavras** (30).

Um teste estatístico de proporções pode ser usado para definir, para 10 ouvintes, qual a porcentagem significativamente distinta de zero para determinado nível de significância. Para o nível de significância de 5%, essa porcentagem é de pelo menos 30%, marcados em negrito acima. Esse critério pode ser usado para indicar palavras proeminentes na fala e seu grau. Por exemplo, a palavra "belo" tem grau de proeminência mais elevado que a palavra "seguida", por ter sido percebida como proeminente por 90% dos ouvintes contra 60% no segundo caso.

No mesmo trecho, pode-se solicitar a tarefa de assinalar as palavras que antecedem imediatamente uma fronteira. Essa tarefa é distinta da de assinalar proeminência. Podemos exemplificar isso no caso da palavra "destes" acima, tida como proeminente por 30% dos ouvintes e, num teste de marcação de fronteiras, assinalada como precedendo uma fronteira por um percentual de ouvintes bem superior a 30%. Apesar dessas diferenças, uma mesma palavra pode desencadear no ouvinte a percepção das duas funções: se uma palavra está em fronteira, sua probabilidade de ser percebida como proeminente é maior do que aquela de uma palavra proeminente ser percebida como antecedendo uma fronteira. Certamente porque uma palavra



Scanned with CamScanner

que antecede uma fronteira chama a atenção do ouvinte de alguma maneira. De qualquer modo, convém ressalvar: uma palavra proeminente pode também se encontrar em posição de fronteira prosódica.

A tarefa em que se pede para o ouvinte delimitar os trechos de fala por fronteiras deve ser privilegiada quando se quer determinar os grupos acentuais pela via da percepção. Recomenda-se o uso de ouvintes leigos na área de pesquisa em fonética, bem como o respaldo dos resultados por análise estatística inferencial[3], conforme sugerido pelo exemplo acima do teste de proporções. A instrução que pode ser passada aos ouvintes pode prescindir da palavra "fronteira", podendo se solicitar algo assim: "Delimite no texto a sua frente, pelo uso de barras, os agrupamentos de palavras que o falante está realizando".

Além da investigação da percepção das duas funções básicas da prosódia, é possível avaliar aspectos mais gerais ligados a diferentes estilos de elocução como: fala lida, fala narrada, fala de telejornalista, sermão religioso, discurso político, entre tantos outros. O pesquisador pode não apenas mensurar os parâmetros prosódicos que se modificam com esses diferentes estilos, usando os procedimentos da seção anterior, como também pode montar testes de percepção para investigar quão distintos são esses estilos, quanto é possível identificá-los para além do conteúdo linguístico veiculado em cada um desses estilos, pois também envolvem uso de vocabulário específico, sintaxe eventualmente específica (por exemplo, o estilo de manchete na fala telejornalística, a parataxe e o maior número de hesitações de algumas narrativas).

Os testes aqui discutidos dizem respeito à prosódia de maneira geral, uma vez que a percepção opera de forma holística, sem separar ritmo de entoação. O que o pesquisador pode fazer é relacionar os resultados de testes de percepção com aspectos temporais (ritmo da fala) ou melódicos (entoação da fala).

O teste que examina diferenças perceptivas é o **teste de discriminação**. Consiste na escuta de dois trechos de som em sequência e em responder imediatamente após escutar uma pergunta simples

[3] É a parte da estatística fundamentada na teoria da probabilidade para inferir o quão consistentes, reprodutíveis são diferenças entre médias, desvios-padrão entre grupos distintos de dados, entre várias outras possibilidades.

sobre a diferença entre os trechos, normalmente associada a graus que variam segundo uma escala assinalada por palavras que indicam gradação. Esse tipo de teste é feito em plataforma específica como a fornecida pelo Praat ou outros *softwares* que elaboram testes de percepção para os quais se podem manipular diversos elementos como aleatorização dos trechos ouvidos entre os ouvintes, tempo entre um trecho e outro, tempo para dar a resposta, ações para dar a resposta (resposta audível, aperto de um botão no teclado ou disposto num espaço na tela), medida de *tempo de reação*, que é o tempo entre o fim da apresentação do estímulo até a ação requerida pelo ouvinte, entre outras possibilidades.

Considere um exemplo em que se avalia a diferença prosódica entre dois estilos de elocução: fala lida e fala de telejornalista. A diferença ou grau de discriminação entre ambos os estilos pode ser avaliada pelos ouvintes que se submetem ao teste a partir de uma instrução como: "Quão diferentes são os modos de falar dos trechos separados pelo tom puro que acabou de escutar?" Observe justamente que o termo "modo de falar" engloba tanto ritmo quanto entoação da fala. Afinal, o uso de termos técnicos seria inapropriado para ouvintes leigos.

Após a escuta dos trechos, separados por um tom puro (para deixar clara a separação entre os dois trechos de fala), o ouvinte clica numa escala na tela com as seguintes opções, numa escala de cinco níveis: "muito distintos", "pouco distintos", "difícil de avaliar", "pouco semelhantes", "muito semelhantes", escala que pode ser equiparada respectivamente aos números inteiros -2, -1, 0, 1 e 2, por exemplo, para a comparação de diferenças acústicas entre os trechos. É justamente ao nível das diferenças acústicas que o pesquisador pode escolher entre avaliar diferenças temporais ao nível silábico (ritmo da fala) ou diferenças melódicas (entoação da fala).

Para ter uma ideia da tarefa de identificação, convidamos o leitor a acessar o *site* e dizer qual dos dois áudios é de um discurso político e qual é fala telejornalística [DelexFemTeste1, DelexFemTeste2]. As respostas estão ao final do capítulo.

Para testes de discriminação de trechos de fala para estudos prosódicos, recomenda-se o uso de trechos com extensões entre 10 e 20



Scanned with CamScanner

segundos. Há duas razões para esse intervalo: por um lado, valores inferiores a 10 segundos não permitem adequada avaliação de um modo de falar que envolva a relação entre alguns grupos acentuais no trecho em questão. Por outro, valores superiores a 20 segundos estão habitualmente além do limiar temporal da *memória de trabalho*, que também inclui a memória das características sonoras de um trecho de fala. A ultrapassagem desse limite prejudica a comparação entre trechos que não estariam mais sendo comparados por seus aspectos sonoros, mas pela retenção de alguma propriedade abstrata fixada após a passagem do limiar temporal.

Na preparação do teste, os trechos selecionados para escuta podem tanto ser apresentados com características sonoras originais ou serem manipulados para eliminação de seu conteúdo lexical, mantendo apenas sua informação prosódica, operação que se chama de *delexicalização*. Isto significa manter exatamente os mesmos valores para os parâmetros de F0, duração, intensidade e qualidade de voz ao longo do enunciado, enquanto o material segmental é distinto, não permitindo identificar o que é dito. Há vantagens em contrastar resultados das duas formas de apresentação, pois é possível inferir a contribuição tanto da prosódia quanto do léxico para a tarefa de discriminação. Há *softwares* disponíveis para realizar a delexicalização como o PURR 🌐.

Além de testes de discriminação, os **testes de identificação** ou **classificação** possibilitam saber se o ouvinte é capaz de identificar um estímulo dentro de um conjunto fechado ou aberto de possibilidades. Se o conjunto for fechado, o teste se denomina *teste de classificação de escolha forçada*. Por exemplo, no caso de identificação de estilos de elocução entre três possibilidades: sermão religioso, discurso político ou fala telejornalística, o ouvinte é convidado a ouvir um estímulo e escolher imediatamente depois entre uma dessas três possibilidades. No exemplo citado, é praticamente impossível a escolha de trechos em que o vocabulário não dê pistas sobre o estilo. Sendo assim, a delexicalização se faz obrigatória para avaliar se, de fato, há uma especificidade prosódica de cada um dos estilos. Se uma das opções de resposta permite que o ouvinte diga que não é nenhum dos estilos propostos, tem-se um teste de classificação de escolha não forçada que permitiria avaliar a ambiguidade de determinados estímulos. Recomenda-se a leitura de material sobre testes psicolinguísticos para alguns cuidados experimentais como: procedimentos de aleatorização de estímulos, uso de estímulos distratores que dificultem ao ouvinte o uso de estratégias enviesadas para dar sua resposta, número de estímulos para que o ouvinte não canse, além de procedimentos estatísticos para avaliar a coerência de respostas do mesmo ouvinte ou entre ouvintes, bem como levar em conta a possibilidade de uso excessivo de falsos alarmes, para acertar respondendo da mesma maneira.

Há outras funções de natureza prosódica que podem ser investigadas a partir de procedimentos idênticos ou similares aos apresentados nesta seção. Alguns deles serão discutidos na seção sobre prosódia e expressividade do capítulo 5. Além da observação e descrição de funções ligadas ao ritmo da fala, ou seja a seus padrões temporais, é possível também modelar esses padrões a partir de algoritmos fundamentados em teorias de organização temporal da fala, como discutiremos no capítulo 6.

🌐 Quanto ao teste de identificação, o arquivo de áudio no *site*, DelexFemTeste1, é da fala de um discurso político e o arquivo de áudio DelexFemTeste2, da fala de uma telejornalista.



Scanned with CamScanner

CAPÍTULO 4

# Entoação da fala

A entoação pode ser definida como a organização na cadeia da fala de padrões de variação de graves e agudos ao longo dos enunciados. A percepção dessa dimensão de variação é chamada de *pitch*. Na literatura sobre entoação, costumam-se distinguir dois elementos básicos que organizam a entoação da fala: o **tom de fronteira** e o **acento de *pitch***.

Os tons de fronteira contribuem para segmentar a cadeia de fala em enunciados e sintagmas entoacionais, unidades apresentadas no capítulo 2. Os acentos de *pitch* desempenham a função básica de proeminência de uma palavra ou palavra fonológica. A conjugação de tons de fronteira e acentos de *pitch* contribui, juntamente com o ritmo da fala, para assinalar atitudes, expressões de afeto, distinções ilocucionárias como pergunta e asserção, além de outras. O estudo experimental da entoação da fala se concentra essencialmente no exame do contorno ou curva de F0 ao longo de um enunciado, também chamada de *curva entoacional*. É, portanto, um componente da organização da fala de longo termo, excluindo elementos que envolvem a variação de F0 de domínio mais restrito como **tom** e **micromelodia**.

O tom se refere ao uso da F0 para contrastar diferentes valores de *pitch* pela indicação de itens lexicais e gramaticais distintos em *línguas tonais* como mandarim, vietnamita, cantonês e iorubá. Línguas tonais são aquelas que variam o *pitch* numa mesma sequência de fonemas para mudar o significado da palavra, como nos exemplos seguintes em mandarim: mā (mãe), com tom alto, e mǎ (cavalo), com tom descendente-ascendente de um total de quatro tons.



Scanned with CamScanner

A micromelodia se refere ao efeito local de um segmento acústico habitualmente do tamanho de um fonema sobre a curva entoacional. Observe na figura 4.1, no trecho lido 4.1 por uma falante paulista de cerca de 20 anos, o aumento inicial do valor de F0 logo após a oclusiva surda /t/ (em "vento") e a fricativa surda /s/ (em "sul") pela posição das duas primeiras flechas e a diminuição local do valor de F0 quando da realização da africada [dʒ] (em "discutiam"), assinalada pela terceira flecha. Como a pressão no trato vocal é maior quando produzimos consoantes não vozeadas como as do exemplo, ocorre um aumento local da vibração das pregas vocais, acarretando um maior valor de F0. Da mesma forma, a pressão cai quando produzimos consoantes vozeadas como a africada sonora do exemplo, fazendo as pregas vocais vibrarem mais devagar e acarretando um valor menor de F0. Em geral, por conta de termos aprendido a associar, durante o processo de aquisição da linguagem, essas modificações com efeitos não entoacionais, mas apenas fisiológicos, não escutamos essas alterações, que, por isso, são chamadas de efeitos micromelódicos.

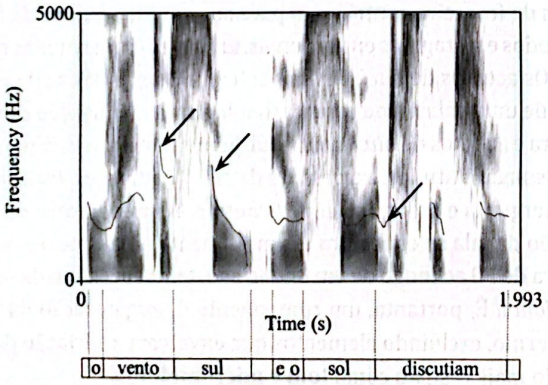


**Figura 4.1:** Ilustração de efeito micromelódico em três locais do trecho "O vento sul e o sol discutiam [...]" de falante paulista jovem.

## 4.1 Tom de fronteira

A figura 4.2 ilustra a curva de F0 de trecho da fala de um locutor da Central Brasileira de Notícias [CBN], que diz: "Ele era uma vítima da democracia. No dia seguinte [...]".

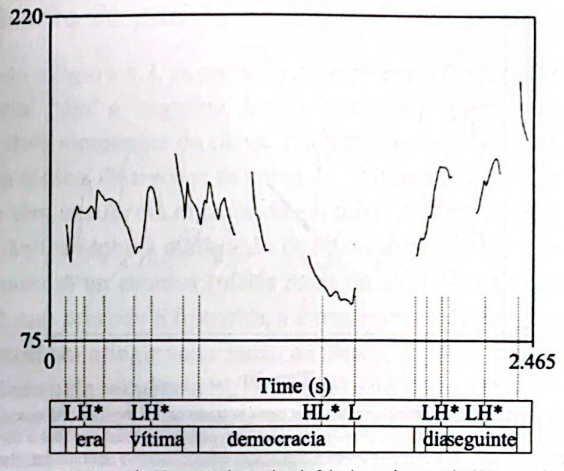


**Figura 4.2:** Curva de F0 em Hz de trecho da fala de um locutor da CBN, que diz: "Ele era uma vítima da democracia. No dia seguinte [...]"

Na figura, foram marcadas apenas as palavras proeminentes e que precedem uma fronteira prosódica. A curva de F0 correspondente ao trecho falado vem acompanhada de símbolos que marcam tons de fronteira (L[low] = tom de fronteira baixo) e acentos de *pitch* (LH* [low-high] = curva ascendente com pico na tônica e HL*[high-low] = curva descendente com vale na tônica). São produtos da pesquisa em entoação da fala, os sistemas de **notação entoacional** desenvolvidos para essas marcações como: o ToBI (*Tone and Break Indices*), concebido primeiramente para o inglês americano e depois adaptado para outras línguas e o DaTO, que simplifica tanto o aprendizado exigido pelo ToBI quanto sua forma de marcação. Para mais detalhes ver no *site* [sistemas de notação entoacional].

Observe que um tom de fronteira, assinalado por L(*low*) pelo fato de a palavra "democracia" soar grave e apresentar nivelamento em nível baixo da curva de F0, foi usado para marcar uma fronteira prosódica, no caso aqui, de enunciado. Observe também que, no enunciado seguinte, que começa com "no dia seguinte", a curva de F0 parte de um ponto mais alto que o nível atingido ao fim do enunciado anterior, marcando assim um reinício da enunciação, mudança que se chama de *reset* (reinicialização) de F0.



Scanned with CamScanner

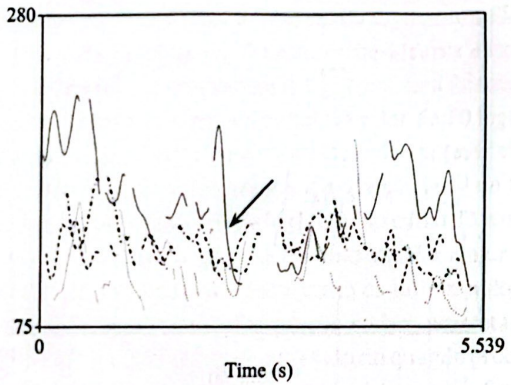


**Figura 4.3:** Curvas de F0 do enunciado "Manuel tinha entrado para o mosteiro há quase um ano, mas ainda não se adaptara à sua maneira de viver." em português brasileiro (linha contínua) e dos enunciados correspondentes traduzidos para o francês (linha pontilhada), alemão (linha tracejada) e português europeu (linha pontilhada-tracejada). A seta aponta o local que precede a conjunção "mas" com tom L para o PB e H para as demais.

O tipo de tom de fronteira pode variar para a mesma função em línguas distintas, o que significa que a uma forma entoacional não corresponde uma única função. A figura 4.3 mostra quatro curvas de F0 para a sentença "Manuel tinha entrado para o mosteiro há quase um ano, mas ainda não se adaptara à sua maneira de viver." [Incipit-PB, Incipit-PE] enunciada em quatro línguas ou variedades do português: português brasileiro e europeu, além de francês ("Manuel était entré au monastère depuis presque un an, mais il ne s'était toujours pas adapté à ce mode de vie.") [Incipit-Français] e alemão ("Manuel war vor fast einem Jahr ins Kloster gekommen, hatte sich aber immer noch nicht an jene Art zu leben angepasst.") [Incipit-German]. A seta na figura assinala a região do tom de fronteira correspondendo aos trechos assinalados por sublinhado em cada língua (um ano, un an, gekommen). Observe que somente o português brasileiro assinala com tom baixo a fronteira não terminal que antecede a conjunção nas quatro línguas. As demais o fazem com tom alto, mas, nas quatro línguas, o respectivo ouvinte nativo entende que o locutor ainda dirá algo. O que causa a percepção de que há algo suspenso em português brasileiro é o pico de F0 que se vê logo antes da fronteira baixa.

## 4.2 Acento de *pitch*

Voltando à figura 4.2, as palavras proeminentes "era", "vítima", "democracia", "dia" e "seguinte" foram marcadas por símbolos que indicam dois elementos da curva: sua forma e seu alinhamento com a sílaba tônica. Os trechos de curva de F0 durante palavras proeminentes têm sua forma organizada em torno da sílaba tônica de forma dinâmica (subida ou descida de F0 ou combinação de ambos os movimentos) ou estática (níveis baixo ou alto). Na palavra "democracia", que precede a fronteira, a curva começa de valor elevado e decresce para atingir valor baixo na tônica. A forma descendente é assinalada pela sequência HL (*high-low*) e o asterisco (*) assinala alinhamento do tom baixo com a tônica. Apesar de diferentes taxas de subida, isto é, da velocidade com que os valores de F0 sobem, e dos diferentes alvos atingidos, as demais palavras do mesmo exemplo têm o tom dinâmico LH (*low-high*) com o alvo alto alinhado com a sílaba tônica de cada palavra. Em português brasileiro, os contornos de subida tendem a ser percebidos como indicativos de continuidade na enunciação, isto é, que o locutor ainda tem o que dizer, enquanto os contornos de descida veiculam habitualmente uma função de terminalidade de uma asserção.

O último acento de *pitch* do sintagma entoacional é denominado na literatura de *acento nuclear*, enquanto aqueles que o antecedem no mesmo sintagma são chamados de *acentos pré-nucleares*.

## 4.3 Alinhamento tonal

Tanto tons de fronteira quanto acentos de *pitch* estão alinhados com as sílabas dos enunciados, especialmente com as sílabas tônicas. Esse alinhamento pode variar, a depender do que se quer expressar, ou da língua.

Observe, na figura 4.4, que as curvas de F0 correspondentes aos acentos de *pitch* sobre as palavras "muita" e "fofoca" do trecho de enunciado 4.4 "É muita colocação de fofoca" estão alinhados com as respectivas sílabas tônicas de formas distintas. Enquanto o pico



Scanned with CamScanner

de F0 se alinha ao final da sílaba "mui-" no primeiro caso (notação LH*), é o vale de F0 que se alinha com a tônica "-fo-"(notação HL*) no segundo caso, a partir de um valor elevado ao longo da sílaba pretônica da palavra "fofoca".

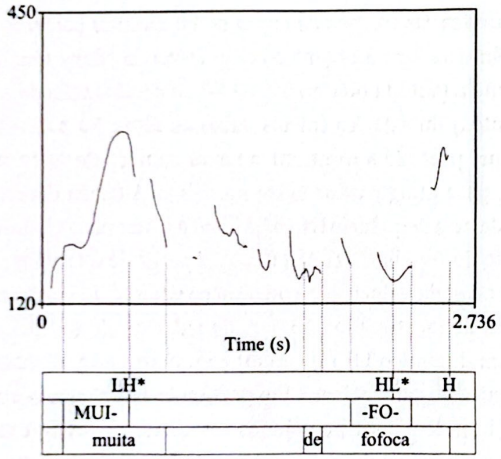


**Figura 4.4:** Curva de F0 em Hz do trecho de enunciado "É muita colocação de fofoca" de locutora adulta em entrevista de programa de rádio.

O trecho termina em tom alto (H) porque a locutora segue dizendo em sua fala o trecho "em cima de um acontecimento que ninguém sabe exatamente o que é."

## 4.4 Aspectos entoacionais de longo termo

Eventos entoacionais como os tons de fronteira e os acentos de *pitch* são, do ponto de vista temporal, eventos de curto termo, pois associados a uma posição no enunciado (tons de fronteira) e a uma palavra ou palavra fonológica (acento de *pitch*). Mas há também eventos entoacionais de longo termo que operam ao longo de várias palavras fonológicas. Esse é o caso do efeito de *declinação*, que pode ser observado tanto pela sucessão dos valores mínimos de F0, que recebe o nome de *baseline* (linha de base), quanto pela sucessão dos valores máximos de F0, que recebe o nome de *topline* (linha de topo).

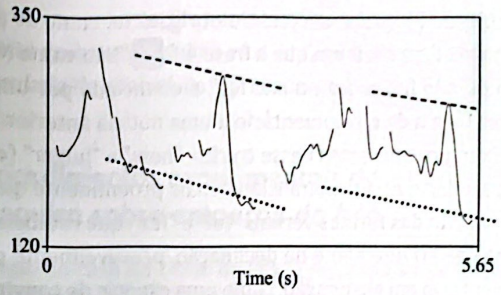


**Figura 4.5:** Curva de F0 em Hz correspondente à frase lida "Manuel tinha entrado para o mosteiro há quase um ano, mas ainda não se adaptara à sua maneira de viver." de locutora paulista jovem. As linhas tracejadas superiores são de declinação *topline* e as linhas pontilhadas inferiores são de declinação *baseline*.

Observe, na figura 4.5, que os valores mínimos de F0 que a curva atinge localmente ao longo do enunciado vão diminuindo até o meio, ponto da palavra "ano", onde há uma fronteira, e retomam um valor ascendente pelo efeito de *reset* (retomada) no início da palavra "mas" e voltam a cair para chegar a um valor na posição final ainda mais baixo do que o encontrado no meio. A mesma diminuição progressiva ocorre com os valores máximos de F0 na *topline*, sendo o último valor de pico de F0 inferior aos valores dos picos que o antecedem.

Embora a declinação seja em grande parte consequência de uma diminuição da pressão subglotal que tem por consequência a redução da intensidade da fala e da taxa de vibração das pregas vocais, muitos autores concordam que a função associada a essa diminuição progressiva de F0 é a de terminalidade.

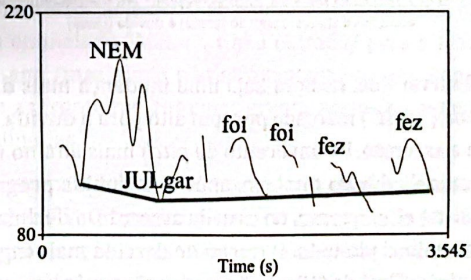


**Figura 4.6:** Curva de F0 em Hz correspondente à frase "Não vamo NEM JULgar se foi ou não foi, se fez ou não fez", de locutor da rádio CBN durante seu programa.



Scanned with CamScanner

A declinação de F0 não é um padrão obrigatório, como se pode ver no exemplo da figura 4.6, em que a frase 4.5 "Não vamo NEM JULgar se foi ou não foi, se fez ou não fez", é enunciada por um locutor de rádio na forma de um comentário a uma notícia anterior. Há duas palavras bem proeminentes nesse trecho, "nem" e "julgar" (essa com acento secundário na primeira sílaba, mais proeminente que a tônica) e a repetição das formas verbais "foi" e "fez", que estabelecem um movimento de F0 que não é de declinação, provavelmente para deixar o comentário em suspensão, como uma espécie de convite a fazer com que o interlocutor não prossiga avaliando a veracidade dos fatos.

Outra mudança global no contorno de F0 pode ser vista com mudanças não linguísticas durante o ato de fala, como no caso de diferentes ilocuções. Isso pode ser visto na figura 4.7 para ilocuções de asserção e de dúvida em francês.

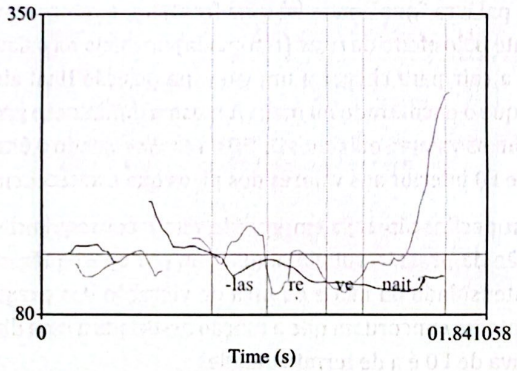


**Figura 4.7:** Curvas de F0 em Hz da sentença "Nicolas revenait" em francês enunciada com duas ilocuções distintas: asserção (preta) e dúvida (cinza).

Pode-se observar que, embora haja uma mudança mais drástica na última sílaba ("-nait") marcada por tom alto para a dúvida e por tom baixo para a asserção, há um acento de *pitch* mais alto no nome "Nicolas" no caso da dúvida também, após uma subida progressiva da curva de F0 até aí, enquanto, no caso da asserção, a declinação se dá ao longo do enunciado todo. O trecho de descida mais íngreme que precede a tônica "las" de "Nicolas" na asserção, após um período em que a curva é interrompida, é um efeito micromelódico do segmento [k] presente na palavra. Ele não é notório na dúvida por conta provavelmente do acento de *pitch* em "-las", que acarreta uma produção hipoarticulada da sílaba "-co-" precedente.

## 4.5 Procedimentos experimentais da pesquisa sobre entoação da fala

Assim como no caso do ritmo da fala, a pesquisa sobre a entoação se aplica essencialmente à observação, descrição e modelamento das curvas de F0 dos enunciados de uma língua para o estudo das funções prosódicas de proeminência e segmentação, bem como para o estudo das consequências para a entoação de mudanças de estilo de elocução, de gênero, de nível de formalidade entre interlocutores, de mudanças no estado emocional, de atitude entre muitas outras situações.

Os procedimentos relatados na seção 3.4 do capítulo 3 podem ser usados para obter palavras proeminentes, grupos acentuais e sintagmas entoacionais que podem ser estudados do ponto de vista da pesquisa em entoação da fala a partir das curvas de F0. Integrar uma descrição das durações silábicas advinda de um estudo rítmico com uma descrição das curvas de F0 é fundamental para investigar como os parâmetros prosódicos se modificam para veicular as funções prosódicas básicas bem como diferentes estilos de elocução, emoções, atitudes, e qualquer outra função, seja linguística, paralinguística ou extralinguística.

A figura 4.8 ilustra as vantagens em se conjugar os perfis prosódicos de dois parâmetros acústicos: F0 e duração normalizada de unidade VV para o enunciado "Manuel tinha entrado/ para o mosteiro/ há quase um ano,/mas ainda não se adaptara/ àquela maneira de viver./", com as fronteiras dos cinco grupos acentuais assinaladas por barras inclinadas.

A curva preta contínua conecta os valores de duração normalizada das unidades V-V. O terceiro pico local corresponde à fronteira que precede a conjunção "mas". Observe que ele se realiza na palavra "ano", assinalada na curva de F0 por um tom de fronteira baixo. O mesmo acontece ao fim do enunciado: tom de fronteira baixo e eleva-



Scanned with CamScanner

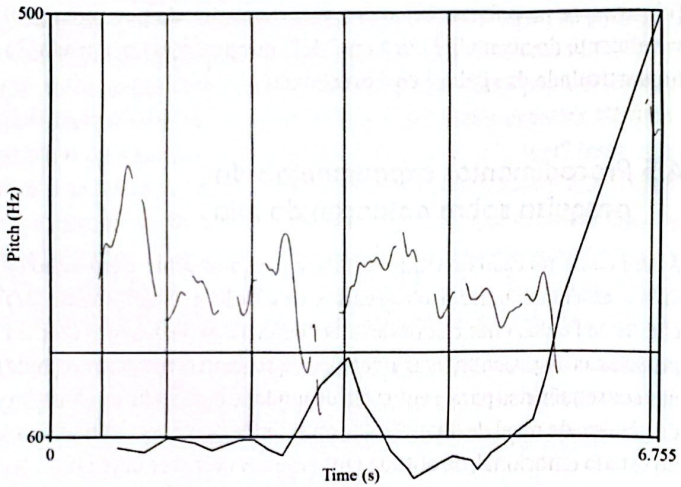


**Figura 4.8:** Curva de F0 (cinza claro) e de duração normalizada de unidades V-V do enunciado "Manuel tinha entrado para o mosteiro há quase um ano, mas ainda não se adaptara/ àquela maneira de viver." por locutora paulista.

do valor de duração normalizada na unidade V-V "-/eR/". O primeiro grupo acentual, terminado por duração normalizada de baixo valor, tem um acento de *pitch* na palavra "Manuel", a personagem principal da história lida. Observe o leitor como é importante observar que tanto F0 quanto duração assinalam as fronteiras depois de "ano" e de "viver", bem como o acento de *pitch* pode ser independente do uso da duração silábica.

### 4.5.1 Obtenção de corpora para estudos prosódicos

Para o estudo de como os parâmetros prosódicos, especialmente rítmicos e melódicos, modificam-se na realização de determinada função, é preciso conceber um experimento controlado para obter situações em que haja contraste entre um trecho em que a função é realizada e um trecho preferencialmente idêntico em que a função não se realize. Assim, pela avaliação do que muda em termos de parâmetros prosódico-acústicos, se descobre como esses parâmetros se organizam para realizar a função em estudo.

Por exemplo, a posição de proeminência no enunciado pode ser manipulada no caso da fala lida a partir do uso de caixa alta no trecho que se deseja ler com foco informacional, como nos casos abaixo. Qualquer leitor fluente saberá como dar mais relevância à palavra em caixa alta. E o estudo, entre os enunciados, do que se modifica entre a situação em que a palavra é proeminente e aquela em que não é, revelará os parâmetros que marcam foco informacional.

A MARIA caiu da escada.
A Maria CAIU da escada.
A Maria caiu da ESCADA.

Se o estudo for de foco contrastivo, será preciso um contexto como nos exemplos abaixo, em que (A) é o contexto e (B), a resposta cujos parâmetros serão comparados com a situação controle, sem foco estreito algum: "A Maria caiu da escada."

(A) A Susana quem caiu da escada?
(B) Não, a MARIA caiu da escada.
(A) A Maria escorregou na escada?
(B) Não, a Maria CAIU da escada.

A obtenção de contrastes dessa natureza em fala não lida é muito difícil, pois será preciso esperar que ocorram a partir de situação espontânea ou guiadas a partir de uma tarefa, cenário ou jogo.

Se a função a ser estudada for a segmentação prosódica na fala lida, ela pode ser manipulada pelo experimentador através de sinais de pontuação ou pelo controle do momento de apresentação de trechos a serem lidos. Já na fala não lida, a segmentação prosódica pode ser manipulada, a depender do tipo de tarefa por diferentes recursos: da sequência ou conteúdo dos quadrinhos, se histórias em quadrinho forem usadas para eliciar fala. No caso de entrevistas, a segmentação pode ser controlada pela forma de interrupção ou condução da entrevista. Já na imitação, pela imposição de modelos a serem imitados com as segmentações que o experimentador precisar variar. Tudo dependerá do que se deseja investigar.



Scanned with CamScanner

Nos diferentes tipos de fala supramencionados, fala lida, fala não lida (entrevista, imitação, obtida de tarefa imposta pelo pesquisador), há aspectos de controle experimental que devem ser considerados para se obter um bom material para estudo prosódico.

A leitura em voz alta pode ser feita apenas a partir de material a ser lido diretamente, sem contexto prévio, como frases isoladas ou textos corridos. Todavia, essa leitura pode ser conduzida a variar em função de público presente ou imaginado. Afinal, a leitura para crianças é diferente da leitura para adultos. Ou se pode imaginar que seja uma leitura para persuadir alguém a comprar algo ou aderir a uma equipe. A simples mudança da natureza da audiência ou do objetivo da leitura modifica a forma como falamos e, portanto, a prosódia.

Se a leitura em voz alta for precedida de uma situação hipotética proposta ao leitor pelo pesquisador, essa leitura pode conduzir a uma mudança de comportamento, como diferentes atitudes, diferentes emoções, diferentes estados de humor.

Se o material a ser obtido para estudo requiser enunciados não lidos, como em jogos ou tarefas a serem feitas, a forma como se instrui o participante pode determinar comportamentos não desejados. Por exemplo, na tarefa chamada de *map task*, um dos participantes deve instruir outro a se dirigir de um ponto A para um ponto B num mapa. Se o mapa tiver elementos muito complexos, a tarefa pode ser tornar complicada e gerar muita hesitação, por exemplo. Esse comportamento hesitativo também pode acontecer, caso o participante tenha problemas de orientação espacial.

A fala imitada a partir de um modelo presente no momento da eliciação ou de um modelo previamente gravado pode servir para revelar ao experimentador os aspectos que mais chamam a atenção do participante do experimento, tanto em situações de imitação imediatamente após o modelo (*close shadowing*) ou após um tempo controlado para se avaliar o que ficou retido na memória de trabalho, por exemplo (*late shadowing*).

Em situações de entrevista, a familiaridade entre experimentador e participante e tema da entrevista são determinantes para a obtenção de um bom material para estudos prosódicos sob pena de se ter material escasso ou com muitas hesitações. Claro, se a hesitação em si não for o tema em estudo.

Descrições a partir de imagens podem ser usadas para se conseguir enunciados controlados minimamente com participantes iletrados. Nesse caso, o mesmo material será passado aos participantes letrados, caso ambos os grupos devam ser comparados para apontar suas diferenças prosódicas.

Para estudos prosódicos de narrativas, por exemplo, pode se usar uma história lida previamente, solicitar a evocação de uma situação vivida recente ou remotamente, usar histórias em quadrinhos para serem narradas, além de outras estratégias de eliciação.

Em todos esses casos, curvas de F0 e padrões duracionais podem ser estudados em diferentes contrastes.

UNIVERSIDADE FEDERAL
Biblioteca Comunitária
S. CARLOS



Scanned with CamScanner

CAPÍTULO 5

# A prosódia da fala e suas interfaces

Este capítulo aborda seis relações da prosódia com outros domínios (sintaxe, significado, discurso, processamento, expressividade e indivíduo), bem como avalia as consequências de seu déficit a partir de lesões encefálicas (disprosódia). O objetivo é dar ao leitor elementos para o estudo das interfaces da prosódia com domínios que requerem conhecimento de disciplinas que não sejam a fonética ou a fonologia.

Assim, o estudo da interface entre prosódia e sintaxe requer do investigador um conhecimento em teorias sintáticas, da mesma forma que o estudo de sua relação com o significado requererá conhecimento em semântica.

O conhecimento de aspectos de análise conversacional e estudos formais de estrutura do discurso fornecem uma base importante para estudos das relações entre prosódia e discurso, da mesma forma que estudos da psicologia experimental sobre afetos fornece uma orientação fundamental para as relações entre prosódia e expressividade.

O reconhecimento de características prosódicas distintas entre os indivíduos tem recentemente despertado o interesse pelo campo da análise forense. Isso exige, por parte do interessado em se aventurar pelo campo, um conhecimento inicial na área de fonética forense.

O estudo do processamento neuronal para a produção e percepção da prosódia requer domínio em áreas como psicolinguística, neurolinguística e psicologia experimental e um olhar atento para as con-



Scanned with CamScanner

sequências linguísticas das lesões neuronais que conduzem a uma adequada investigação da disprosódia.

## 5.1 Prosódia e sintaxe

A questão mais relevante no que diz respeito ao tema desta seção é a da **não congruência** entre prosódia e sintaxe. Tomemos o enunciado 5.1 "Maria canta divinamente bem.", pronunciado com esta segmentação: [Maria canta][divinamente bem]. Ora, do ponto de vista sintático, a primeira ramificação da sentença seria a que estabelece os constituintes $[Maria]_{SN}$ e $[canta\ divinamente\ bem]_{SV}$, em que "Maria" é um sintagma nominal com função de sujeito e "canta divinamente bem" é um sintagma verbal com função de predicado. Por outro lado, a quebra principal no enunciado se dá depois de "canta" para assegurar uma divisão em grupos acentuais de extensão semelhante ("Maria canta" tem 5 sílabas e "divinamente bem" tem 6). Essa tendência à regularidade dos grupos acentuais é uma propriedade prosódica que impõe uma fronteira principal distinta da proposta pela divisão sintática. Há assim uma não coincidência das fronteiras principais entre os constituintes prosódicos e sintáticos, cunhada na literatura pela expressão "não congruência".

Nesse caso, a estrutura sintática da sentença é recuperada pelo ouvinte mediante seu conhecimento da sintaxe do português brasileiro. Em outros casos, a prosódia pode guiar a interpretação sintática de uma sentença. Observe os enunciados abaixo obtidos a partir de distintos grupos acentuais com uma fronteira não terminal forte entre constituintes marcada por "/":

[A ovelha de raça]/[brasileira] (5.2)
[A ovelha]/ [de raça brasileira] (5.3)

No primeiro caso, se trata de uma ovelha de raça não informada nascida no Brasil e, no segundo caso, de uma ovelha que é de raça desenvolvida no Brasil. São os constituintes prosódicos que permitem a escansão adequada da estrutura sintática de cada sentença. Isto é, a prosódia permite a *desambiguação* entre as duas interpretações possíveis. Os limitados recursos da escrita não permitem resolver a distinção. Nesse caso, a prosódia guia a interpretação sintática, e os constituintes sintáticos e prosódicos são consequentemente congruentes. Também o são em casos em que a extensão dos grupos acentuais é semelhante entre constituintes sintáticos, como no caso do enunciado "Maria canta bem", com a seguinte escansão sintática e quebra prosódica: $[Maria]_{SN}$ / $[canta\ bem/]_{SV}$.

Pode-se ver claramente, nos exemplos acima, que escolhemos a função prosódica de segmentação para falar da relação entre prosódia e sintaxe. Mas essa relação também se dá com a função de proeminência. Tomemos o caso do contraste entre foco estreito informacional e foco largo que vimos na introdução.

A realização prosódica de um constituinte em foco largo, como no enunciado 3.3, correspondente à sentença "João comprou [uma casa verde]$_F$", é a palavra "verde", a mais à direita, que recebe o acento nuclear. É mais uma vez o conhecimento do português brasileiro e o fato de que a última palavra não tem o perfil acústico de uma palavra em foco contrastivo (como em "João comprou uma casa [VERDE]$_F$") que permite distinguir esse exemplo como tendo um foco largo e não estreito.

Pela função de segmentação, a prosódia permite marcar outros constituintes sintáticos, como os parentéticos. Em enunciados como o 5.4, "Foi Gustavo, engenheiro famoso.", a oração após a vírgula é parentética, pois dá uma explicação secundária à informação principal dada por "foi Gustavo", em resposta a alguma pergunta precedente.

A figura 5.1 ilustra a curva de F0 de um enunciado da sentença "Foi Gustavo, engenheiro famoso.", destacando os dois constituintes prosódicos. Fica evidente que o trecho parentético tem valor de F0 em nível inferior ao do constituinte precedente. Ao medir os valores de F0, sua média é de 82 Hz durante "Foi Gustavo" e de 75 Hz durante "engenheiro famoso". Além disso, o constituinte parentético tem 55 dB de intensidade média contra 63 dB do constituinte precedente. E mais, sob o aspecto rítmico, a taxa de elocução do primeiro constituinte é de 6,3 sílabas (fonológicas) por segundo, contra 7,5 sílabas/s do segundo constituinte. Sendo assim, o parentético soa mais grave,



Scanned with CamScanner

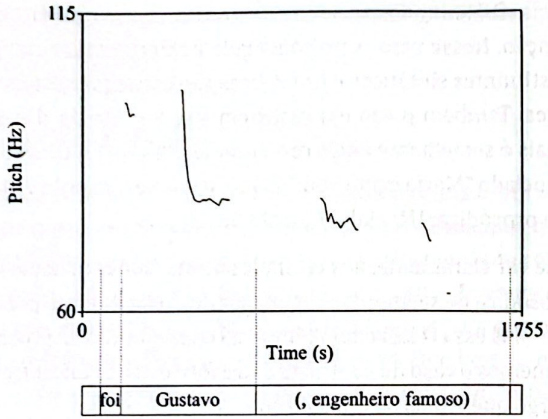


**Figura 5.1:** Curva de F0 da sentença "Foi Gustavo, engenheiro famoso." por um falante paulista de cerca de 50 anos.

é menos intenso e é falado mais rapidamente; através dessas três medidas acústicas em contraste, a prosódia assinala a função parentética do segundo constituinte.

No exemplo de desambiguização entre duas estruturas possíveis a partir da escrita dado acima, fica claro que se, de um lado, a prosódia guia a interpretação da estrutura sintática nos casos de congruência com a sintaxe, ela também é autônoma em relação à sintaxe, por obedecer a princípios rítmicos com o da regularidade de constituintes prosódicos. Em ambos os casos, a prosódia associada à estrutura sintática que o ouvinte recuperou com seu eventual auxílio (no caso de não congruência, o ouvinte deve usar conhecimento sintático de sua língua) guia a interpretação do enunciado.

## 5.2 Prosódia e significado

Os exemplos apresentados na seção precedente evidenciam contrastes de estruturas sintáticas que têm direta implicação semântica, como no caso das duas interpretações possíveis da sentença ambígua "A ovelha de raça brasileira." Essa ambiguidade se dá apenas na escrita, pois, quando a frase é enunciada conforme o contexto, não há margem de dúvida quanto a sua interpretação, como visto acima: a prosódia guia a interpretação correta da frase.

Há casos em que a mesma sequência de palavras tem diferentes funções ilocutórias marcadas tanto na escrita quanto pela prosódia. É o caso da oposição entre ilocução de asserção em "João fez isso." *vs.* questão total em "João fez isso?" As distintas interpretações marcadas pela prosódia assinalam duas ilocuções diferentes e, portanto, atos de fala distintos, diferentes significações. A diferença prosódica nesse contraste se dá primariamente pela diferença de forma ou perfil da curva de F0, como vimos antes. Diversos outros significados para a mesma sequência de palavras podem ser obtidos apenas pelas diferenças de forma prosódica que assinala diferentes ilocuções. Isso fica bem claro nos exemplos usados por Raso (nesta coleção): "João", "vai", "para", "o", "Rio", que pode ser realizado como:

(1) sequência de chamado e de ordem: "João!""Vai para o Rio!",
(2) asserção: "João vai para o Rio.".
(3) confirmação de questão de outrem e asserção: "João?""Vai para o Rio.", entre outras possibilidades.

Em outros casos, como o de distinção entre foco contrastivo e foco largo dos exemplos de "João viu uma moto VERDE." *vs.* "João viu [uma moto verde]$_{F}$.", diferenças de perfil de F0 e alinhamento tonal permitem a adequada interpretação entre os dois enunciados, e, portanto, o acesso aos seus significados distintos.

O conjunto dos parâmetros prosódico-acústicos permite realizar o acento lexical, importante elemento linguístico diferenciador de palavras em português. Esses parâmetros auxiliam na recuperação do significado de enunciados que contêm palavras em contraste quanto ao acento lexical. Um exemplo de contraste dessa natureza é aquele entre o enunciado 5.5 "Não sei se caso de dia." e o enunciado 5.6 "Não sei se casou de dia." em que a distinção entre "caso"['kazu] e "casou" [ka'zo] é realizada tanto pela diferença da qualidade dos fones finais como por diferenças prosódicas para implementação do acento lexical, diferenças ao nível da duração, F0 e intensidade entre as duas sílabas em cada palavra como assinalado na figura 5.2.



Scanned with CamScanner

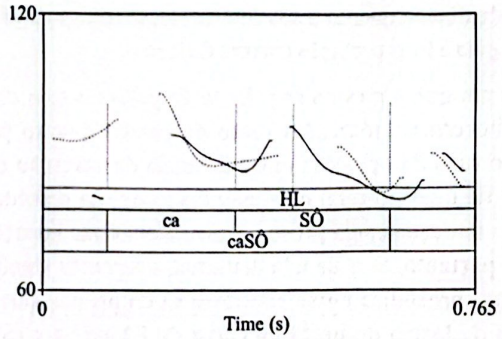


**Figura 5.2:** Curvas de F0 em Hz superpostas dos trechos contendo as palavras em contraste "caso" (linha contínua) e "casou" (linha tracejada, segmentada) num falante de cerca de 50 anos com sotaque paulista.

Observe que há um trecho de curva descendente durante a tônica em cada palavra. Como a tônica de "caso" está na primeira sílaba da palavra, o trecho descendente começa a palavra em "caso" e termina a palavra em "casou" (pronunciada "casô"). Essa diferença contribui para a percepção da posição do acento lexical e, portanto, para distinguir os significados dos dois enunciados. No exemplo dado, se se trata da incerteza sobre o período do dia para o casamento ("não sei se caso de dia") ou se se trata sobre a mesma incerteza, mas a propósito do casamento de um terceiro ("não sei se casou de dia"). É também importante o exame das diferenças de duração e de intensidade entre as duas sílabas, contraste maior na paroxítona. Caso o leitor disponha de prática com o *software* Praat, poderá observar a segmentação das sílabas fonológicas a partir da qual se podem retirar as seguintes medidas respectivas de duração e de intensidade para as duas sílabas de "caso": 259 ms/133 ms e 65 dB/61 dB. Observe: a primeira sílaba dura praticamente o dobro da sílaba átona seguinte, sílaba que é também menos intensa em 4 dB. Quanto às sílabas de "casou", tem-se para a duração: 228 ms/275 ms, aumento de cerca de 20% entre átona e tônica e valores idênticos de 64 dB de intensidade. Embora a avaliação da reprodutibilidade dessas diferenças requeira gravações de repetições desses enunciados em contraste e análise estatística inferencial, as diferenças assinaladas aqui são observadas frequentemente em estudos fonéticos.

## 5.3 Prosódia e discurso

É bem conhecido, na área de análise da conversação, o papel da prosódia na estruturação dos diálogos e conversas entre interlocutores, em que se ressalta como conceito fundamental a **tomada de turno** (*turn taking*).

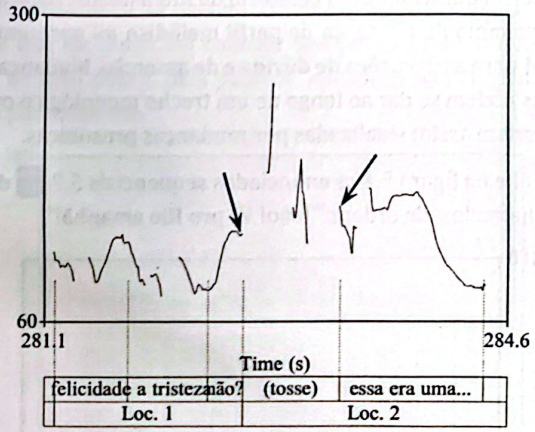


**Figura 5.3:** Curvas de F0 em Hz de trecho de entrevista de programa de TV entre entrevistador e entrevistado. A primeira flecha marca o fim do primeiro turno e a segunda, o início do turno do interlocutor. O trecho entre eles é de tosse do segundo interlocutor.

Essa tomada de turno pode ser cooperativa ou competitiva. A figura 5.3 apresenta o perfil melódico de um exemplo de tomada de turno cooperativa. Nela, o primeiro interlocutor termina seu turno fazendo uma pergunta a seu entrevistado, com a curva de F0 num valor elevado (primeira flecha). Após um trecho de tosse, que afeta a vibração das pregas vocais, e por isso há traços de F0 entre as duas flechas, o segundo interlocutor começa a falar com valor de F0 ligeiramente superior ao do entrevistador, como que continuando seu perfil melódico, algo comum nesse tipo de tomada de turno. De fato, o perfil melódico ascendente, compatível com o de uma "pergunta", do primeiro interlocutor "convida" o segundo a falar. Assim, a prosódia possibilita as trocas de turno no diálogo.

Nas tomadas de turno competitivas, o interlocutor que quer "roubar" o turno do participante na conversa utiliza meios prosódicos



Scanned with CamScanner

para tal: aumenta a intensidade de sua fala, muitas vezes sobre a do outro, e continua falando para esperar que o outro ceda a vez.

Outros aspectos do discurso dizem respeito a como a prosódia veicula diferentes ilocuções, assinalando-as por diferentes **formas prosódicas**. O termo "forma" é entendido como um contorno melódico, um contorno duracional, um contorno de intensidade, habitualmente. Um exemplo de diferença de perfil melódico foi apresentado na seção 4.4 para as ilocuções de dúvida e de asserção. Mudanças ilocucionárias podem se dar ao longo de um trecho monológico ou dialógico, e seriam assim sinalizadas por mudanças prosódicas.

Acompanhe na figura 5.4 os enunciados sequenciais 5.7 de ilocução de chamado e de ordem: "João! Vá pro Rio amanhã!".

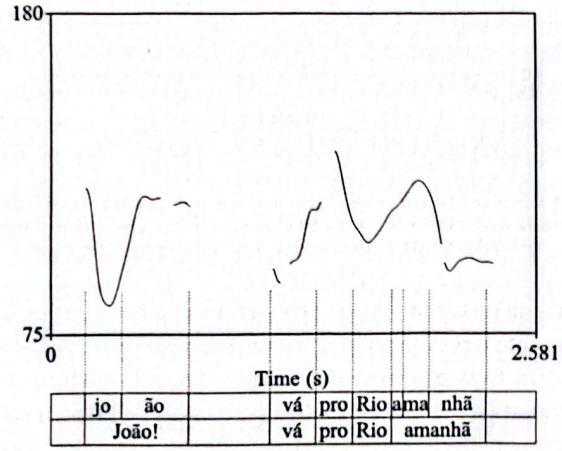


**Figura 5.4:** Curvas de F0 em Hz dos enunciados de chamado ("João!") e de ordem ("Vá pro Rio amanhã!"), de falante de cerca de 50 anos da região paulista.

Nessa sequência, observe o trecho alto de F0 mantido durante a tônica de "João", o trecho ascendente durante o verbo que estabelece o tipo de ordem ("vá") e a descida ao final de "amanhã" iniciada após ponto alto na pretônica "-ma-". Esses perfis conduzem o ouvinte à correta interpretação das funções das respectivas ilocuções, perfis associados a uma maior duração da sílaba final de "João" no chamado e a uma maior intensidade no verbo "vá" na ordem.

A função prosódica básica de segmentação permite, pelo agenciamento de fronteiras terminais e não terminais, organizar o discurso. A figura 5.5 mostra o contorno de F0 do trecho 5.8 "e com os cadernos todos... encader/" de fala espontânea de falante mineiro adulto. Após a palavra "todos" há uma fronteira não terminal. Observe como a curva de F0 sobe para um valor elevado na sílaba tônica e desce na pós-tônica da palavra. O valor elevado na tônica indica ao ouvinte que ele ainda vai concluir seu pensamento, o que começa depois de uma pausa, pela palavra "encadernados", aqui mostrada apenas parcialmente.

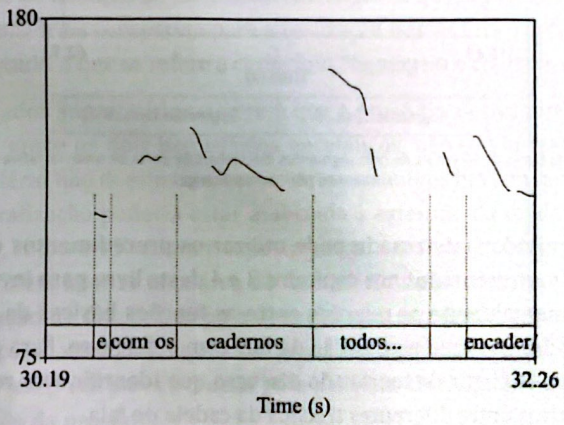


**Figura 5.5:** Curva de F0 em Hz do trecho "e com os cadernos todos... encader/" de fala espontânea de falante mineiro adulto.

A prosódia permite, ainda no discurso, organizar unidades temáticas, a partir de pausas longas entre os temas e de um reinício bem marcado da curva de F0, como no exemplo 5.9 da narrativa de uma lenda amazônica, a do uirapuru, pela atriz Camila Pitanga em um cd comercial. No segmento discursivo 8, ela finaliza com a expressão "ele então fugiu" com valor de F0 final de cerca de 125 Hz, em trecho de sua narrativa que descreve a atitude do índio perante a decepção das índias com sua aparência. Observe que o segmento discursivo 9 continua a narrativa marcado por um reinício de valor de F0 bem mais alto, cerca de 240 Hz. O trecho narrado conta o interesse das índias pelo canto belíssimo de outro índio. A variação



Scanned with CamScanner

desse parâmetro é bem maior nesse novo trecho, contribuindo para a compreensão de que se trata de novo tema.

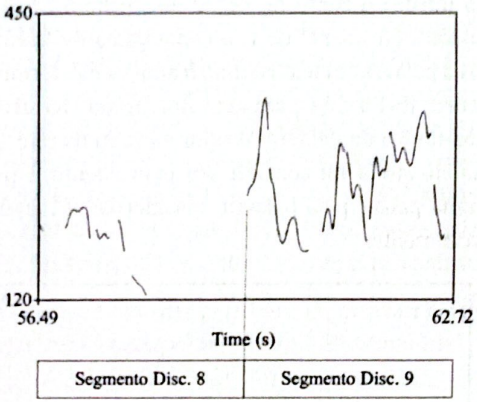


**Figura 5.6:** Curva de F0 em Hz de dois segmentos discursivos de parta de uma narrativa de lenda amazônica por Camila Pitanga.

O pesquisador interessado pode utilizar os procedimentos experimentais apresentados nos capítulos 3 e 4 deste livro para investigar experimentalmente as relações entre as funções básicas da prosódia ligadas ao ritmo e entoação da fala com o discurso. Para tanto é importante partir de teorias do discurso que identifiquem relações discursivas entre diferentes trechos da cadeia de fala.

## 5.4 Prosódia e processamento

Sob a rubrica de prosódia e processamento se subscrevem os estudos que investigam as regiões de nosso *encéfalo* envolvidas na produção e na percepção da prosódia. O encéfalo reúne o cérebro, o cerebelo e o tronco encefálico. Essa rubrica costuma separar a **prosódia afetiva**, voltada ao processamento do afeto, da **prosódia linguística**, restrita às funções unicamente linguísticas. Assim, as três principais questões que norteiam a literatura com relação ao processamento da prosódia podem assim ser formuladas:

- ✓ Há divisão de trabalho entre os hemisférios esquerdo e direito para o processamento da prosódia afetiva *vs.* prosódia linguística?
- ✓ Quais regiões do encéfalo estão envolvidas na produção e percepção da prosódia e o que elas processam em termos de parâmetros prosódico-acústicos?
- ✓ Produção e percepção da prosódia são tratadas por regiões distintas do encéfalo?

Sobre a questão da **lateralização**, ou seja, a divisão de trabalho entre os hemisférios esquerdo e direito, estudos usando a técnica da ressonância magnética funcional mostraram que apenas as *áreas neocorticais*[1] estão envolvidas com funções prosódicas como as de **foco estreito** e de ilocução de enunciado (asserção *vs.* questão) e que ambos os hemisférios concorrem para a realização dessas funções e, em geral, de tudo o que se refere a raciocínio, linguagem e controle motor.

Os estudos supracitados sugerem que a prosódia afetiva também se divide entre os dois hemisférios encefálicos, não sendo restrita ao hemisfério não dominante, como vários trabalhos haviam mostrado. A lateralização poderia estar associada à extensão da unidade prosódica, como evidenciaram experimentos: a realização de foco numa sílaba afeta apenas uma área do hemisfério dominante, enquanto a realização de funções como a ilocução, que se relacionam ao enunciado como um todo, requisitam áreas dos dois hemisférios. Assim, a escala temporal parece crucial para a divisão de trabalho no processamento da prosódia.

Alguns autores defendem a hipótese, sustentada por um grande número de resultados experimentais, de que essa escala temporal distinta é gerada no sistema nervoso não apenas de forma distribuída, mas também de forma diferenciada para funções distintas, sendo **cerebelo** e **gânglios basais** as regiões responsáveis pelo processamento.

Usando diferentes técnicas de observação, estudos mostraram que o cerebelo controla a atividade de ciclos silábicos com taxa entre 4 e 6 sílabas por segundo. Um deles, usando a técnica da resposta temporal pós-estímulo (PST, na sigla em inglês), evidenciou uma atividade neuronal amplificada para porções de aumento de energia do sinal,

[1] As áreas neocorticais são as camadas do cérebro de 2 a 4mm de espessura, que constituem a massa cinzenta, que contém neurônios e fibras não mielinizadas.



Scanned with CamScanner

acompanhada da supressão de atividade para porções em que há diminuição da energia do sinal. Essa atividade acompanha, assim, a variação de intensidade de fala típica da produção silábica.

Outras áreas encefálicas acompanham essa atividade, como as redes de neurônios da porção ascendente da via auditiva que chega ao **girus temporal superior (GTS)**[2], bem como a porção que chega aos **sulcos temporais superiores** abaixo do GTS, que respondem especialmente a oscilações de intensidade da ordem de 4 Hz, como visto acima.

Quanto à percepção da prosódia, o estudo de Steinhaeur e colaboradores usou a técnica de POTENCIAIS EVOCADOS POR EVENTOS (ERP, da sigla em inglês, *Event-Related Potential*) para mostrar que fronteiras prosódicas ouvidas estão associadas a trechos de aumento da amplitude do potencial evocado que foi nomeado de CPS (*Closure Positive Shift*), independentemente da presença ou não de pausa silenciosa, pois há resposta neuronal quando há alongamento de uma vogal, também um fenômeno marcador de fronteira. Como a fronteira é marcada pelo concurso dos três parâmetros prosódicos clássicos: duração silábica, F0 e intensidade, o estudo não permite avaliar qual desses parâmetros elicia maior resposta nas redes neuronais. O que se sabe até o momento é que, como um todo, a percepção da fala ativa tanto áreas corticais bilaterais temporais, como vimos acima, quanto áreas parietais.

Quanto à lateralização de diferentes parâmetros acústicos, há uma predominância, em indivíduos destros, do processamento temporal no hemisfério esquerdo, enquanto o processamento espectral (de frequência) ativa majoritariamente áreas do hemisfério direito, o que também é confirmado por estudos em indivíduos lesionados, seja no hemisfério esquerdo, seja no hemisfério direito, com os primeiros perdendo capacidade de processamento temporal. Além de o hemisfério esquerdo processar preferencialmente unidades temporais, ainda são as de curto termo (da ordem do intervalo do tamanho do fonema) que são privilegiadas. Para unidades temporais mais extensas, há um concurso de áreas encefálicas dos dois hemisférios.

Se o que foi visto até agora diz respeito ao papel de diferentes estruturas para privilegiar produção ou percepção da fala, se sabe que as **áreas parietal** e **frontal** do hemisfério esquerdo dominante estão envolvidas na integração entre produção e percepção da fala.

[2] Que contém a **área de Wernicke** e a **área do córtex auditivo primário.**

## 5.5 Disprosódia

Lesões nas áreas encefálicas responsáveis pelo processamento prosódico geram déficits tanto na produção quanto na percepção dos componentes melódicos e rítmicos da fala. O termo **disprosódia** refere-se às consequências, para a prosódia, de problemas no controle e/ou na execução dos sistemas responsáveis pela produção da fala: respiratório, laríngeo e supralaríngeo, bem como a problemas relacionados à percepção da prosódia. A depender da extensão e do local de uma lesão encefálica, esses problemas de natureza prosódica podem afetar um aspecto específico (apenas a duração ou apenas a F0, por exemplo, ou ainda apenas a duração de unidades de maior extensão temporal, mas não as de menor). Podem ainda afetar somente a produção, mas não a percepção ou vice-versa. Há autores que mencionam também problemas com a prosódia afetiva, mas não linguística, mas quanto a essa separação estanque há controvérsias.

Observe o exemplo da figura 5.7, que ilustra o contorno de F0 de trechos segmentados da produção da sentença 5.10 "Hoje ele vive atrás das grades do zoológico da cidade." por uma falante adulta operada de fissura do palato.

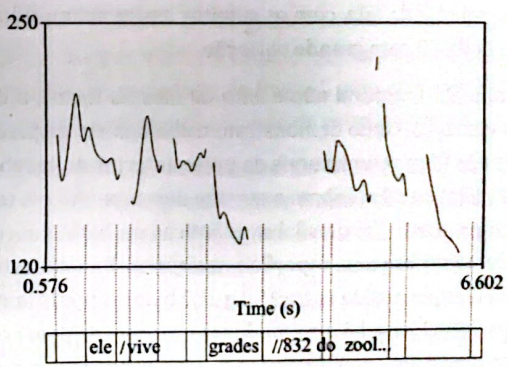


**Figura 5.7:** Curva de F0 em Hz da produção da sentença "Hoje ele vive atrás das grades do zoológico da cidade." por uma falante adulta operada de fissura do palato.



Scanned with CamScanner

Tal problema gera uma dificuldade na extensão da fase expiratória, o que tem por consequência tomadas de ar mais frequentes, que segmentam a fala em unidades bem menores com relação ao que faria um falante sem patologia de fala, chamado em estudos experimentais de *falante controle*.

De fato, no exemplo acima, a falante controle não produz pausa alguma ao ler o trecho, enquanto a falante com fissura palatina introduziu pausas silenciosas entre um pronome e o verbo "viver" (de 181 ms), bem como entre a palavra "grades" e seu sintagma preposicional complemento (pausa longa de 832 ms).

Hoje ele / (181 ms) vive atrás das grades // (832 ms) do zoológico da cidade.

Nesse caso, trata-se assim de uma falha na organização temporal da fala por motivos respiratórios. A curva de F0 é compatível com a de um falante controle, embora pareça-se com o perfil melódico de uma pessoa que faz um ditado, justamente por conta de produzir segmentos de fala curtos.

Estudos com sujeitos lesionados em diferentes regiões encefálicas têm mostrado que acometimentos no hemisfério esquerdo associados ou não a acometimentos no cerebelo costumam afetar majoritariamente o controle temporal, portanto, o ritmo da fala, enquanto lesões do hemisfério direito provocam alterações na produção e percepção da melodia da fala, com os sujeitos assim acometidos produzindo curvas de F0 sem grande variação.

O processamento temporal não é feito da mesma forma, a depender da taxa de elocução, como demonstram trabalhos em disprosódia. De fato, lesões nas fibras comissurais da parte anterior do cérebro fazem com que os sujeitos não tenham o mesmo desempenho em tarefas repetitivas como produção de sílabas idênticas ou batida sucessiva de um dedo da mão sobre uma superfície horizontal. Para taxas de repetição na faixa compreendida entre 150 e 200 batidas por minuto (BPM), isto é, entre 2,5 e 3,3 Hz, há uma mudança na forma de processamento, de maneira que, acima de 200 BPM, o sujeito lesado tem o mesmo desempenho em relação ao grupo controle, enquanto para as taxas abaixo de 150 BPM, seu desempenho é muito inferior ao grupo controle. Isso demonstra que a região cerebral é responsável por um controle analítico, isto é, para extensões temporais maiores, sendo os movimentos rápidos controlados por outra região encefálica. Esse outro tipo de processamento que trata repetições rápidas de um movimento é holístico, como a produção de uma oscilação em que se controla não cada período em si, mas a própria frequência de oscilação.

Doenças neurodegenerativas, como a doença de Parkinson (DP) e a doença de Huntington (DH), são também caracterizadas por alterações prosódicas. Na grande maioria dos casos, apenas a produção da prosódia, e não sua percepção, são afetadas na DP, enquanto na DH tanto produção quanto percepção da prosódia são afetadas. No caso de Parkinson, a fala apresenta pausas mais longas com pequenas variações de F0 (soa mais monótona) e de nível de intensidade e ainda com taxa de elocução pouco variável. Esses problemas são gerados tanto por **disartria** quanto por **apraxia**, problemas presentes nos dois tipos de doença.

A disartria é uma falha ao nível muscular na articulação da fala, resultando numa fala não somente mais lenta, mas hipoarticulada. Já a apraxia é uma falha ao nível do controle da articulação, isto é, a falha ocorre antes, já ao nível encefálico.

Do ponto de vista experimental, os estudos de disprosódia requerem uma cuidadosa avaliação clínica sobre a localização e extensão das lesões, bem como a seleção de um grupo controle, isto é, de pessoas saudáveis que tenham as mesmas características sociolinguísticas (idade, sexo, escolaridade, variedade de língua), para que comparações adequadas quanto à preservação ou não da prosódia possam ser feitas.

Os aspectos de produção que devem ser investigados são todos aqueles apresentados nos capítulos sobre ritmo e entoação da fala. Assim, deverão ser comparados os seguintes descritores ou variáveis entre grupos controle e experimental (o grupo dos pacientes):

- ✓ taxas de articulação e elocução;
- ✓ realização das pausas silenciosas (frequência de pausas, duração, posição em relação a constituintes prosódicos observados no grupo controle);



Scanned with CamScanner

- ✓ duração de sílabas átonas e tônicas
- ✓ realização de proeminências e fronteiras;
- ✓ aspectos globais de melodia (declinação de F0, realização global de funções como atitudes e emoções);
- ✓ laringalização (frequência e em que segmentos e trechos de fala ocorrem), entre outros.

Quanto à percepção, testes de percepção da prosódia devem ser montados para investigar a capacidade do grupo experimental em perceber as funções básicas de proeminência e segmentação da fala, bem como os aspectos rítmicos e melódicos tanto linguísticos quanto afetivos.

Somente um exame cuidadoso das eventuais diferenças entre os grupos controle e experimental pode lançar luz sobre o grau de disprosódia dos indivíduos em estudo.

## 5.6 Prosódia e expressividade

A prosódia também exerce funções não linguísticas que veiculam um grande número de aspectos expressivos da fala. Basta pensar nas diferentes interpretações de atores mesmo na mesma peça teatral, ou nas diferentes formas de narrar uma mesma história, de declamar um mesmo poema, que se relacionam com as diferentes formas de interpretar uma canção, campo de investigação da prosódia musical.

A figura 5.8 representa as curvas de F0 de dois momentos da interpretação da história de Pedro Malasartes por Camila Pitanga representando a fala de duas personagens, uma com tom de voz mais grave (esquerda), enunciado 5.11 🌐, e outra, com tom mais agudo (direita), enunciado 5.12 🌐. Essas diferentes sensações são uma consequência dos valores mais altos de F0 no segundo trecho. De fato, nesse segundo trecho, o valor mínimo de F0 é de 305 Hz, o máximo de 497 Hz e a média de 396 Hz. Já no primeiro trecho, o valor mínimo de F0 é de 217 Hz, o máximo de 392 Hz e a média de 296 Hz. Observe que a atriz consegue estabelecer uma diferença de cerca de 100 Hz entre esses descritores estatísticos para representar as diferentes personagens.

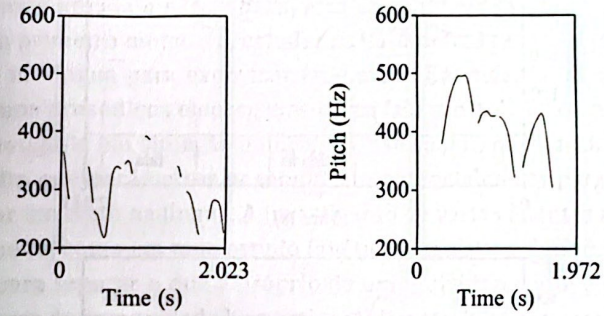


**Figura 5.8:** Curvas de F0 em Hz de dois momentos da interpretação da história de Pedro Malasartes narrada por Camila Pitanga representando a fala de duas personagens distintas.

Além da interpretação de personagens, a modificação da estrutura prosódica do fluxo da fala gera um grande número de efeitos expressivos com o fim de agradar, causar determinada impressão no ouvinte, como ocorre na declamação de poemas, leitura de obras literárias, narração de histórias para crianças.

Além desses aspectos lúdicos, a expressividade manifesta pela prosódia permite chamar a atenção do ouvinte para uma locução telejornalística ou esportiva ou um sermão religioso. Pode ainda procurar persuadir o interlocutor a comprar um produto, aderir a uma campanha, votar num candidato, ser convencido de uma ação política como num pronunciamento de Ministro ou de Presidente à nação. Todas essas diferentes manifestações prosódicas requerem mudanças temporais, melódicas e de intensidade da fala e mesmo de qualidade de voz.

Diferentes atitudes com relação a uma proposição ou a um interlocutor podem ser manifestas pela prosódia, como vimos no capítulo 4, no caso de contraste entre dúvida e asserção. Essas atitudes podem incluir ainda situações como a do sarcasmo.

Observe na figura 5.9 o contraste entre dois enunciados provindos da sentença "Você tem que tomar cuidado com quem fala.", entre dois cenários para eliciar uma situação de leitura neutra, descontextuali-



Scanned with CamScanner

Scanned with CamScanner



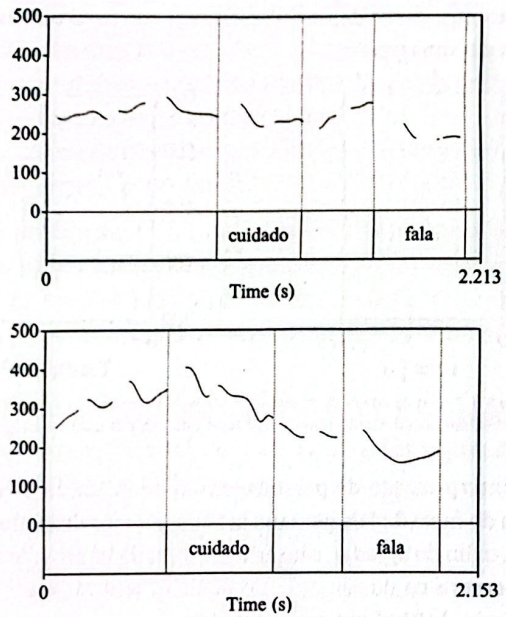


**Figura 5.9:** Curvas de F0 em Hz da sentença "Você tem que tomar cuidado com quem fala." enunciada de forma neutra (superior) e sarcástica (inferior) por locutora paulista.

zada, enunciado 5.13, e outro em que o sarcasmo é a atitude que emerge no falante, enunciado 5.14. Observe que a curva de F0 tem valores superiores em todo o enunciado com sarcasmo (entre 200 e 400 Hz) com relação ao enunciado neutro (entre 200 e 300 Hz). O desvio-padrão de F0 no enunciado neutro é de apenas 26 Hz contra 69 Hz na fala sarcástica. Além disso, a intensidade global durante a palavra "cuidado" passa da forma neutra para a sarcástica de 65 a 72 dB e, durante a palavra "fala", de 56 a 69 dB, configurando um volume superior, imprimindo uma sensação de força, de energia ao sarcasmo. A duração da palavra "cuidado", tema central da sentença, é maior também na fala sarcástica, como se depreende do trecho da palavra nos dois enunciados na figura, uma vez que a escala na abscissa é praticamente a mesma (cerca de 2,2 segundos).

Observe o leitor que o contraste entre uma situação neutra e outra experimental, que elicia a situação que se deseja investigar, é fundamental para entender as mudanças prosódicas produzidas para veicular o sarcasmo. Evidentemente, a situação ilustrada acima é atuada e não natural, mas somente assim se consegue, num primeiro momento, estudar os parâmetros prosódicos usados na língua para expressar esse afeto. Entendendo os mecanismos prosódicos empregados nessa fala controlada, é possível investigá-la em situação natural, pois se saberá que parâmetros medir, que parâmetros se modificam potencialmente para assinalar um afeto na língua. A investigação de vários falantes a esse respeito, como em todo estudo fonético-linguístico, é fundamental para separar o que é próprio de um indivíduo daquilo que é recurso da comunidade linguística. Assim também se procede no estudo da emoção na fala.

Para alguns psicólogos experimentais, a emoção é um dos possíveis estados do afeto e precisa ser distinguida de outros estados, como o humor, as atitudes e as disposições do afeto. A emoção pode ser distinguida desses outros afetos por pelo menos quatro aspectos principais:

(1) menor duração do estado ou processo afetivo em relação ao evento desencadeador (algo entre meio e 4 segundos para alguns autores);
(2) maior intensidade de respostas corpóreas (fisiológicas, motoras e de tendência à ação);
(3) maior sincronismo em relação ao evento ou objeto desencadeador;
(4) relação direta com uma avaliação cognitiva do estado ou processo afetivo, sobretudo no que diz respeito ao impacto comportamental.

No que diz respeito ao estudo da emoção, quatro perspectivas teóricas podem ser identificadas, e podem ainda ser estendidas ao estudo da fala expressiva em geral. Resumindo, essas perspectivas identificam:

(1) a emoção como causadora de uma resposta generalizada que engloba a face e a voz, uma vez que ela é fruto da movimentação de estruturas corpóreas, perspectiva que decorre das ideias de Darwin em seu livro *As emoções nos homens e nos animais*;
(2) a emoção como resultado dessa resposta corpórea inata, como se fosse uma espécie de sensação que fica depois que o corpo

reage ao estímulo causador, perspectiva promovida pelo psicólogo William James;

(3) a emoção como fruto de uma avaliação cognitiva que ocorre dinamicamente à medida que se manifesta, perspectiva conhecida como *cognitivista*;

(4) a emoção considerada como construto eminentemente social numa perspectiva *socioconstrutivista*.

Embora haja um movimento de rejeição ao radicalismo da última proposta, não se pode esquecer que as emoções podem ser como que mascaradas ou atenuadas por questões sociais, como a evocação da regra de conduta "homens não choram" deixa transparecer.

As duas primeiras perspectivas estão enraizadas numa referência biológica que sugere que a emoção deve ser entendida por seu valor para a sobrevivência da espécie. Por esse motivo as duas perspectivas defendem a universalidade da emoção. Enquanto a pesquisa darwinista ressalta o aspecto expressivo da emoção, a perspectiva fundamentada no trabalho de James ressalta o papel de sua experiência. Para ambas as perspectivas, no entanto, as emoções são adaptações ambientais com função de sobrevivência da espécie.

Quanto à *universalidade da emoção*, estudos de emoções em outras espécies, bem como o estudo de expressões faciais, recolheram evidências para um conjunto de seis emoções básicas que definiriam um sistema composicional que explicaria outras emoções. Essas emoções são a alegria, a tristeza, o medo, o desprezo, a raiva e a surpresa, embora haja pequenas variações quanto ao número e natureza dessas emoções nessa lista mais comum.

A perspectiva cognitivista se constrói a partir da consideração de que as emoções dependem da avaliação de um sujeito que percebe, noção muito próxima da ideia de que as emoções são tendências à ação: a avaliação informa o organismo de características específicas do ambiente e o predispõe a agir sobre essas características, que é uma forma de adaptação ao meio. Essa perspectiva é importante para os estudos prosódicos porque, em algumas abordagens, reconhece um conjunto de primitivos ou dimensões de avaliação como prazer ou **valência**, **atenção** ou domínio do que ocorre e grau de **ativação**.

Enquanto a valência se refere à avaliação de algo como agradável ou desagradável em diferentes graus, a atenção se refere à avaliação do grau de adesão ou rejeição de quem se emociona com o evento disparador da emoção. A ativação, por sua vez, avalia o quão calmo ou agitado está quem está sob aquela emoção.

Tanto quanto para a expressão facial, os sujeitos são muito precisos em avaliar a expressão vocal dos indivíduos diretamente ou a partir desses primitivos. Quem não reconheceria, na alegria, seu aspecto prazeroso (valência positiva), em que o indivíduo parece aderir à situação que lhe causa a alegria (atenção elevada) e que o coloca num determinado estado de agitação (ativação alta)?

A avaliação de primitivos emocionais relacionados à expressão do afeto faz-se em sessões experimentais com número variável de ouvintes, normalmente com duração que evite cansaço ao avaliar os enunciados. Essa avaliação pode ser feita com o uso de *escalas semânticas diferenciais*, um conjunto pequeno de palavras organizadas numa escala subjetiva crescente identificada entre dois extremos com apenas as palavras que marcam esses extremos ou com palavras para marcar cada grau entre eles. Nesse último caso, a escala é conhecida como *escala de Likert*. O número de graus varia de quatro a nove. Evitar um número que se assemelhe àquele de sistemas de atribuição de notas na cultura específica pode ser uma boa maneira de evitar a transferência de hábitos em atribuir essas notas.

A importância da expressão vocal é desde cedo reconhecida por Darwin, que a equiparava à expressão facial. De fato, a conclusão mais consistente da pesquisa sobre aspectos prosódicos da emoção na fala é a de que emoções com maior valor de ativação correlacionam-se com maiores valores da média e da variância da frequência fundamental, que separam emoções de menor valor de ativação das de maior valor dessa dimensão. Aspectos ainda negligenciados na literatura são os papéis da qualidade vocal e da intensidade relativa no estudo prosódico das emoções na fala.

Tomemos como exemplo de variação prosódica na emoção um caso de depoimento real para o filme-documentário *Jogo de cena*, de Eduardo Coutinho. Selecionamos aqui trechos de depoimentos de



Scanned with CamScanner

duas mulheres que contam experiências de vida. Uma delas fala alegremente de festas que fazia com o filho, enquanto outra fala tristemente da perda de seu filho.

As curvas de F0 da figura 5.10, de extensão praticamente igual, possibilitam o exame desse parâmetro acústico nas duas situações. A expressão da alegria, pela primeira mulher, tem variação e valores médios semelhantes ao longo do depoimento, com amplitude de variação de 132 a 396 Hz com desvio-padrão de 54 Hz. A expressão de tristeza, até a região central, tem amplitude de variação de 109 a 311 Hz com desvio-padrão de apenas 30 Hz. Em seguida, a expressão de tristeza se exacerba, numa espécie de abandono, e a amplitude de variação passa a uma variação de 178 a 424 Hz, com desvio-padrão de 55 Hz. Observe que o segundo padrão se assemelha aos valores do parâmetro para a expressão da alegria na outra mulher.

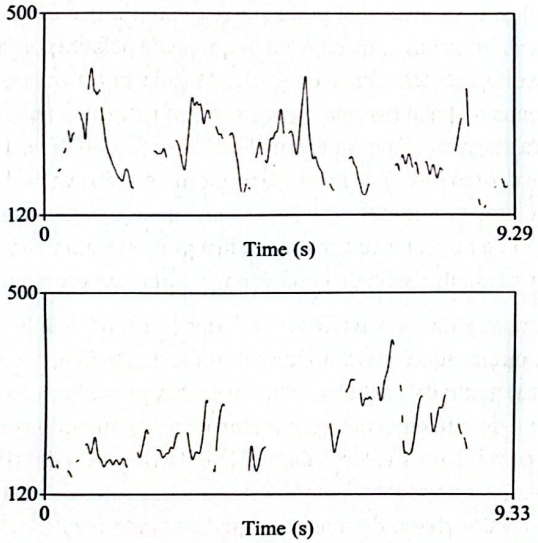


**Figura 5.10:** Curvas de F0 em Hz de trechos de depoimento de duas locutoras cariocas. A curva superior é uma fala em que se expressa alegria. A curva inferior é uma fala com expressão de tristeza. Observe no meio da curva inferior um aumento dos valores e da variação de F0.

Quanto à ênfase espectral para um corte de frequência baixa de 400 Hz, seu valor é de 2 dB para a alegria e passa de 2,4 a 1,2 dB nas duas fases da tristeza. Assim, o esforço vocal é maior na primeira parte do depoimento triste, demonstrando uma maior dificuldade, uma maior tensão vocal ao começar o relato. A alegria demanda menos esforço vocal global. Menos ainda quando a mulher quase que libera suas emoções, relaxando a tensão vocal na expressão de maior tristeza na segunda parte do relato.

Quanto à taxa de elocução, se ela é de 6,1 sílabas (fonológicas) por segundo no depoimento de alegria; na tristeza, a taxa começa com esse mesmo valor no primeiro trecho e diminui para 5,2 sílabas por segundo no segundo trecho, numa lentificação típica da expressão da tristeza.

Observe que, na pesquisa sobre expressividade, é possível conjugar uma análise obtida de dados mais controlados, em que a fala expressiva é eliciada pelo experimentador, com uma análise da fala expressiva espontânea, a partir do que se aprende no primeiro caso. O contraste entre uma situação neutra e outra expressiva é fundamental, mas só é possível no caso da fala eliciada, controlada. De fato, esse contraste é raríssimo na fala espontânea, uma vez que é pouco provável encontrar-se uma enunciação neutra com mesmo contexto fonético da enunciação expressiva.

Os procedimentos de exame de parâmetros temporais e melódicos vistos nos capítulos 3 e 4 devem ser empregados na investigação da fala expressiva, que ganha ainda quando se adiciona aspectos de qualidade de voz, relacionados à tensão ou relaxamento da laringe.

A expressividade da fala normalmente é estudada nos aspectos comuns a vários indivíduos, com o objetivo de descobrir as características gerais de determinado afeto. Por exemplo, a alegria se relaciona a uma agitação que causa maior tensão nas pregas vocais e maiores excursões dos articuladores de fala, com consequências para os parâmetros acústicos melódicos e temporais: média e desvio-padrão mais elevados de F0, taxa de elocução modificada (mais alta, em geral), maior intensidade. A tristeza, por outro lado, causa a diminuição dos mesmos parâmetros. No entanto, considerando esses aspectos, há espaço para a expressão própria de um indivíduo. Por vezes, a forma de dar ênfase, a forma como usa as pausas e sua extensão, as



Scanned with CamScanner

mudanças em sua qualidade de voz ou mesmo a forma como ri, como chora, ao mesmo tempo em que fala. Reconhecer um indivíduo pela maneira como se expressa também é interesse da prosódia.

## 5.7 Prosódia e indivíduo

Há dois interesses centrais no estudo da relação entre prosódia e indivíduo. Um deles é a questão da *verificação de locutor* e o outro, a *identificação de locutor*. Um exemplo de uso típico da tarefa de verificação de locutor é a chave de segurança vocal em que as características fonéticas de um indivíduo são comparadas com as de outros indivíduos previamente gravados para acesso a um local ou a dados. Nesse caso, o indivíduo colabora com o dispositivo que verifica sua voz para garantir seu reconhecimento como uma das pessoas que pode ter autorizado o acesso.

Na tarefa de identificação de locutor, mais complexa, o que se quer é reconhecer um indivíduo, em princípio, num conjunto aberto, isto é, sem que se saiba se ele faz parte de um conjunto de indivíduos cujas características fonéticas foram colhidas anteriormente. É a tarefa típica da área de fonética forense, que aplica os conhecimentos fonéticos ao campo forense, procurando reconhecer se a fala de um indivíduo sob investigação criminal num material colhido numa investigação policial é a de um dos potenciais suspeitos. A diferença da tarefa anterior é que, excetuando-se a situação de tentativa de acesso fraudulento a dados ou a um local seguro por chave vocal, o indivíduo que se quer reconhecer, o criminoso, pode não estar entre os suspeitos.

Em nosso caso, interessa apontar quais possíveis características prosódicas podem permitir reconhecer um indivíduo entre outros. E podemos começar afirmando que essa tarefa deve ser feita em conjunto com a análise de características segmentais, isto é, da investigação cuidadosa de como são realizados os diferentes fones da fala, pois mesmo a forma de se pronunciar um simples "r" pode ser determinante para o reconhecimento de um indivíduo. Assim, é um conjunto amplo de parâmetros acústicos, segmentais e prosódicos, que pode ajudar a afunilar o conjunto inicial de suspeitos, cuja voz "bata" com a do investigado por um crime.

Um dos aspectos importantes nesse processo é procurar identificar, numa avaliação inicial, o que parece diferente, idiossincrático na fala de um indivíduo. Foquemos nos aspectos prosódicos, por conta do escopo deste livro. Embora a F0 média seja ainda um parâmetro usado para o reconhecimento de um indivíduo, há que se tomarem certos cuidados. Se o valor médio for próximo à media geral da F0 de determinada comunidade linguística, é delicado apontar algo tão comum como um elemento que apontaria alguma individualização. Há que considerar também que, sob tensão, na situação de crime, o indivíduo pode mudar muito sua fala, como se viu na seção anterior. Assim, mesmo que não se encontrem entre os suspeitos características melódicas similares, não se pode descartar a possibilidade de o investigado ser um deles com base na comparação de duas situações de fala distintas: a recolhida na investigação policial e a recolhida na entrevista com o policial com o consentimento do investigado ou através de seu advogado.

A figura 5.11 mostra, em duas situações, as curvas de F0 de uma mesma pessoa que estava sob investigação em dois momentos distintos, gravações colhidas por peritos de dois órgãos públicos. O áudio não é disponibilizado para o livro tendo em vista o caráter do material. A mediana de F0 da curva superior, de uma das gravações, é de 141 Hz contra 123 Hz no segundo momento, curva inferior. O valor mínimo nos dois trechos é muito semelhante, no entanto, de cerca de 110 Hz.

A variação das curvas é muito diferente, pois, no momento que corresponde à curva inferior, o indivíduo está falando com mais calma. Assim, é preciso ter cuidado em concluir muito rapidamente a distinção ou não de duas amostras de fala.

Alguns dos parâmetros prosódicos que têm sido usados para a identificação de locutor são: média de F0, valores máximo e mínimo de F0, distância do valor máximo de F0 com relação ao início da vogal, duração silábica e variação de energia do sinal de fala. Pesquisa recente mostra que parâmetros rítmicos como taxa de elocução e articulação, a forma como alguns indivíduos prolongam ou encurtam



Scanned with CamScanner

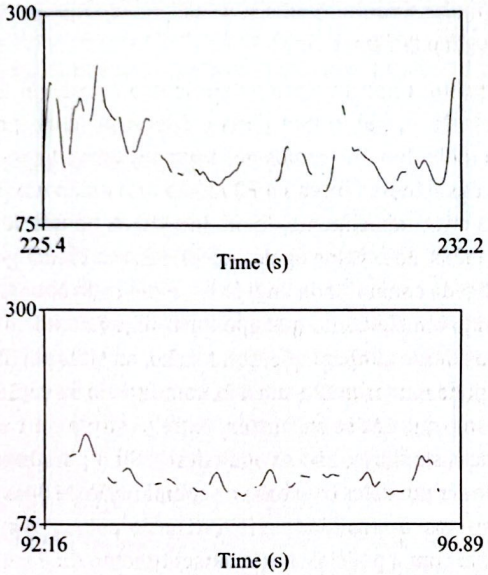


**Figura 5.11:** Curvas de F0 em Hz de um mesmo indivíduo dando depoimento a um perito em análise forense. Observe que a do painel inferior está num nível de valores inferior e varia menos.

sílabas, a forma de fazer ênfases, entre outras, podem ser promissores para a identificação de falante. Outro aspecto importante é a observação das disfluências, que são de mais difícil controle pelo indivíduo. É assim o mais tipicamente idiossincrático e o menos prosodicamente modificável que deve ser buscado pelo profissional que faz a análise forense.

Além da comparação estatisticamente fundamentada entre parâmetros prosódicos, são empregadas técnicas avançadas de aprendizado de máquina (*machine learning techniques*), que consideram um grande número de parâmetros acústicos para a identificação de locutor. Todos esses meios devem considerar importantes aspectos da diferença de qualidade acústica do material sob comparação. O material de fala *questionado*, obtido durante a investigação sem o conhecimento do indivíduo, é, hoje em dia, normalmente provindo de conversas por meio de telefone celular em ambiente ruidoso, enquanto o material de *referência*, gravado com equipamento de qualidade com o consentimento do indivíduo, tem uma relação sinal-ruído bem mais elevada. Há distorções no cálculo dos parâmetros acústicos do material questionado que devem sempre ser levadas em conta na análise forense e a escuta atenta dos dois tipos de amostra é fase fundamental para uma análise fonético-forense de qualidade.

Some-se a isso a necessária conjugação de pistas fornecidas pelo maior número possível de parâmetros prosódicos (e segmentais) para compor um relatório técnico que suporte, numa escala composta normalmente de nove graus, a probabilidade de que as amostras de fala analisadas provenham do mesmo locutor.



Scanned with CamScanner

CAPÍTULO 6

# Teorias e modelos

As teorias e modelos da prosódia procuram fornecer uma explicação da forma como nós produzimos e percebemos a prosódia da fala. Sua primeira preocupação é propor como base uma separação fundamental entre a produção e a percepção da prosódia e entre a produção e a percepção de consoantes e vogais. A produção é considerada como tendo por base a sílaba, a menor unidade prosódica, a ela se subordinando a produção dos segmentos vocálicos e consonantais.

Por teoria, entendemos um conjunto de asserções testáveis sobre variáveis observadas e, por modelo, um modo de gerar valores a partir dessas variáveis. Assim, uma teoria sobre o ritmo da fala propõe explicar como as durações silábicas seriam produzidas e percebidas num determinado contexto comunicacional.

Há teorias da prosódia tanto de cunho fonológico quanto fonético. Como este livro diz respeito à prosódia da fala, se concentrando sobre a investigação de natureza fonética, estão excluídas de seu escopo teorias como a da fonologia entoacional de Ladd e a fonologia prosódica de Nespor e Vogel. Assim, privilegiamos a apresentação de teorias fonéticas de produção/percepção da prosódia que são importantes para entender a separação fundamental entre essa e os segmentos.

## 6.1 Teorias prosódicas pautadas em princípios biológicos

A teoria *frame/content* (quadro/conteúdo) de Macneilage e colaboradores e a teoria de Shattuck-Hufnagel e colegas, *slots/fillers* (más-



Scanned with CamScanner

caras/preenchimentos), apresentadas por volta dos anos 1980, são semelhantes na forma de conceber essa separação.

Nessas teorias, a propriedade organizacional da fala é caracterizada pelo movimento cíclico e alternante de abertura e fechamento da boca, movimento que especifica, a cada ciclo, a sílaba (*frame/slot*), derivada de ciclos de oscilação mandibular presentes nos humanos desde o balbucio. Essa oscilação teria se desenvolvido quando oscilações mandibulares relacionadas à ingestão (atos de mastigar, sugar e lamber) tomaram significado comunicativo.

A partir do movimento de abertura e fechamento mandibulares, deve ter se desenvolvido filogeneticamente, por necessidade comunicativa, a independente diferenciação de vogais e consoantes (*content/filler*). Essa mesma diferenciação se repete de forma paulatina, mas na escala do desenvolvimento humano, durante a aquisição da linguagem pela criança.

Investigações conduzidas a partir de engenhoso aparato experimental que avaliou taxa de sucção por recém-nascidos revelam que a prosódia da língua materna é reconhecida desde o primeiro instante do nascimento, o que sugere a aquisição da prosódia durante o período intrauterino. Vários estudos mostram que, de fato, isso é possível desde a formação do sistema auditivo em torno da 26ª semana de gestação. O feto ouve as vibrações externas que lhe chegam através do corpo da mãe. Se pensarmos num paralelo bem simples, é como escutarmos alguém falando do outro lado de uma parede: escutamos o modo como ele fala, mas não o conteúdo, isto é, escutamos aspectos de ritmo e entoação, mas não identificamos as vogais e consoantes. Esse quadro teórico, sustentado por um número cada vez maior de estudos, propõe que o recém-nascido percebe a prosódia da fala que ouviu durante a gestação. Aliado ao aspecto eminentemente oscilatório da mandíbula, o bebê tem tudo o que é preciso para produzir, imitar, reproduzir a fala de seu entorno.

Esses modelos teóricos experimentalmente fundamentados propõem assim que a produção e a percepção da prosódia precedem, na esfera comunicativa, a produção e a percepção dos segmentos. É como se a criança dispusesse de uma máscara oscilatória que vai aos poucos estendendo da repetição da mesma sílaba para a variegação, a inserção de segmentos distintos. Começa associando vogais e consoantes articulatoriamente próximas (como /di/./gu/) e consoantes facilmente visíveis como as bilabiais com vogais abertas relacionadas ao choro ou grito (como /ma/) para depois explorar outras combinações. Essas tentativas vão se construindo com o aumento dos grupos acentuais formados por uma única sílaba, em seguida por duas, para compor um contraste acentual mínimo, até a produção de grupos acentuais maiores, num processo que vai até a adolescência, quando da aquisição do ritmo e entoação do adulto.

Além de tratar da questão da separação dos componentes segmental e suprassegmental da fala, as teorias da prosódia se debruçam sobre como se dá a organização estritamente melódica ou estritamente rítmica da cadeia da fala. No caso da melodia ou entoação, as teorias propõem modelos lineares ou superposicionais da entoação. Os *modelos lineares* da entoação a conceibem como uma sucessão não composicional de tons altos (H) e baixos (L) ou combinados (e.g., L+H, H+L). São modelos habitualmente de caráter fonológico que pouco se ocupam de como a curva de F0 é realizada. O tipo de notação oriundo de modelo linear foi usado no capítulo 4 para ilustrar acentos de *pitch* e tons de fronteira.

## 6.2 Modelos de geração da melodia da fala

Tendo em vista o interesse por um modelo que privilegie a fala em seus aspectos de produção e percepção, convém nos referir aqui primordialmente aos *modelos composicionais*, que são modelos de geração da curva de F0. O mais conhecido deles foi desenvolvido por Fujisaki com a colaboração de seu então aluno Hirose nos anos 1980.

A figura 6.1 traz uma representação esquemática do modelo de Fujisaki, que propõe a geração, em escala logarítmica, da curva de F0 a partir da soma de três componentes: a frequência de base ou valor mínimo $F_m$; o componente de sintagma entoacional (*phrase component*), equivalente a um mecanismo de reinício do movimento de F0 que demarca esse tipo de sintagma, e o componente de acento

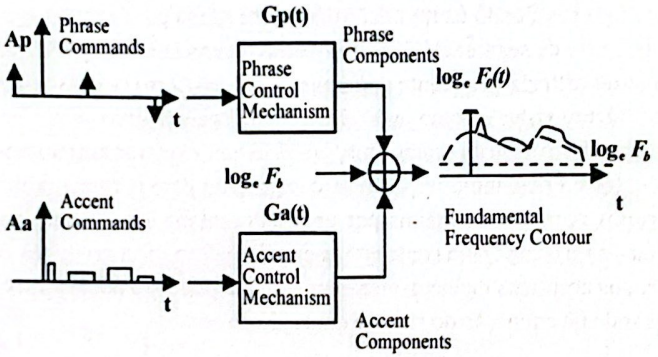


**Figura 6.1:** Modelo de geração de F0 de Fujisaki. Reproduzido com autorização de Hirose.

(*accent component*), que realiza cada acento de *pitch*. O resultado da adição ou superposição desses três componentes pode ser visto à direita da figura: a linha tracejada horizontal é o valor mínimo ou $F_b$, a linha tracejada com curva convexa e extensa define os limites dos sintagmas entoacionais a partir dos comandos de sintagma (*phrase commands*) e a linha cheia é obtida ao final, aplicando-se os comandos de acento (*accent commands*). O modelo de Fujisaki é usado extensivamente até os dias de hoje. Talvez a maior dificuldade seja conceber um método para automaticamente gerar os dois tipos de comando numa determinada língua. Quando isso é feito, é possível compreender os princípios que regem a segmentação e a realização de proeminências melódicas na língua. Daí a importância de se ter um modelo: permitir análises automáticas mais simplificadas, eliminando parte da variação, aplicando-se o mesmo procedimento a todo um *corpus*.

O modelo PENTA de Xu também é um modelo de geração da curva de F0 gerada numa escala logarítmica, cujos trechos são alinhados com unidades prosódicas como sílabas ou grupos acentuais. Embora não seja um modelo aditivo como o de Fujisaki, as diversas funções linguísticas afetam de forma independente a forma geral da curva final levando em conta alvos estáticos ou níveis e alvos dinâmicos ou inclinações. Exemplo do tipo de curva de F0 gerada pelo modelo PENTA pode ser visto na figura 6.2.

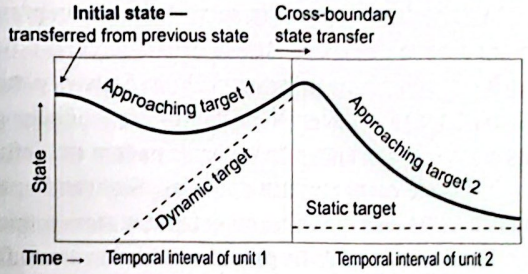


**Figura 6.2:** Modelo de aproximação de alvo do modelo PENTA de Xu. Reproduzido com autorização de Xu.

Observe que a curva de F0 se aproxima do primeiro alvo dinâmico (*dynamic target*) ao longo da unidade 1 e depois se aproxima do alvo estático (*static target*), já na unidade 2. Alvos dinâmicos são aqueles que assinalam subida ou descida de F0, enquanto alvos estáticos assinalam um valor fixo a ser atingido. Essas unidades são os domínios para a realização de determinada função prosódica da língua. No caso da implementação do acento de *pitch*, podem ser sílabas ou grupos acentuais. Se à função de acento de *pitch* se superpõe a de segmentação, o modelo PENTA modifica localmente a curva de F0, de forma a implementar o tom de fronteira.

Por conta de as funções linguísticas nesse modelo serem implementadas de forma paralela, isto é, a realização de uma é independente da realização da outra e o efeito de uma se superpõe ao das outras, o modelo se classifica como superposicional. Outros modelos superposicionais usam técnicas de aprendizado automáticas para a separação de componentes melódicos associados a diferentes unidades prosódicas, como o modelo de Morlec, que considera o nível global do enunciado, do sintagma entoacional e do grupo acentual. Cada componente da melodia em cada nível é gerado a partir de um aprendizado prévio a partir de dados de fala lida. A ideia é a mesma da do modelo de Fujisaki, mas nesses outros modelos as unidades prosódicas são definidas *a priori* e impostas para o aprendizado do modelo.

Por se pautarem em princípios articulatórios de produção da curva de F0, tanto o modelo de Fujisaki quanto o de Xu acabam por conceber modelos fisiologicamente plausíveis que se coadunam com



Scanned with CamScanner

a percepção da melodia. Ambos se servem de uma representação logarítmica da curva de F0 que aponta para a percepção da sensação de *pitch*, como vimos no primeiro capítulo. Embora o modelo de Fujisaki não imponha de antemão os limites das unidades prosódicas, seus comandos de acento e de sintagma podem ser definidos de forma alinhada com essas mesmas unidades, alcançando plausibilidade linguística. Os dois modelos, no entanto, tratam a duração de maneira secundária, atrelada às próprias unidades prosódicas. Daí a importância de modelos que tratem exclusivamente da geração da duração de unidades prosódicas.

## 6.3 Modelos de geração da duração e do ritmo da fala

Os primeiros modelos de geração da duração foram concebidos nos anos 1970 no contexto da *síntese da fala a partir do texto*, isto é, a geração automática da fala a partir do texto escrito, modelo que gera os valores para todos os parâmetros acústicos de determinado enunciado. O principal deles é o modelo de Klatt, que gera a duração de cada segmento isomórfico ao fonema sem preocupação com aspectos prosódicos de forma integrada. Cada fone tem sua duração obtida a partir da modificação de uma duração intrínseca, modificação que depende de fatores contextuais, como proximidade de uma fronteira ou segmento vozeado, entre outros.

Seguiram a mesma linha outros modelos de geração da duração para sistemas de síntese da fala como os modelos de van Santen e os de Bartkova e Sorin, modelos também segmentais, na medida em que não concebem uma duração de unidade prosódica como a sílaba e, portanto, não levam em conta o ritmo da fala. É justamente esse tipo de falha que os modelos prosódicos de geração do ritmo da fala vêm corrigir, pois levam em conta a duração silábica concebendo sua modificação ao longo do enunciado, de forma a realizar funções como a segmentação e a realização de uma prominência na cadeia de fala.

O modelo de Byrd e Saltzman, denominado $\pi$-gesture, implementa um alongamento local em fim de grupos acentuais a partir da concepção de um "gesto" $\pi$, dotado de força variável, representando, assim, diferentes níveis de proeminência ou de marcação de fronteira. No modelo desses autores, as fronteiras mais fortes provocam efeitos de alongamento maior nos gestos segmentais sob o domínio do gesto prosódico. O modelo deles considera ainda diferentes formas do gesto prosódico, que produz efeitos não lineares variados sobre os gestos segmentais. Os autores consideram a modificação da duração silábica somente numa janela específica, excluindo qualquer modificação que atuasse, por exemplo, ao longo de todo um grupo acentual. Essa é uma diferença importante com relação ao modelo de geração do ritmo da fala de Barbosa, um modelo de osciladores acoplados.

Esse modelo integra uma teoria dinâmica do sistema rítmico que subjaz à fala, pressupondo três níveis de *acoplamento* (influência mútua) em três escalas temporais distintas com propriedades biológicas e linguisticamente plausíveis. Em primeiro lugar, o modelo postula que o ritmo da fala advém do acoplamento entre dois comportamentos opostos: uma tendência à estruturação ou diferenciação de duração e uma tendência à regularidade silábica. A primeira tendência é implementada no modelo pelo acoplamento entre a informação sintática local e o **oscilador acentual**. A segunda tendência, por sua vez, é implementada pela presença de um **oscilador silábico,** que, na ausência do oscilador acentual, oscila de forma periódica. No modelo, as duas tendências contrárias (porque a primeira tende a diferenciar durações e a segunda tende a igualá-las) operam em escalas temporais distintas, uma escala da ordem da magnitude do grupo acentual, outra, da magnitude da sílaba. O oscilador silábico do modelo tem seus ciclos ancorados na sequência de inícios das vogais, propriedade crucial para revelar o ritmo da fala, como vimos no capítulo 3.

O modelo de osciladores acoplados gera padrões complexos de duração silábica como resultado da modificação dos ciclos do oscilador silábico pela influência do oscilador acentual. Esses padrões são compatíveis com aqueles de duração normalizada encontrados em enunciados naturais do português brasileiro, como os padrões vistos no capítulo 3. A taxa de elocução é uma propriedade intrínseca do modelo, dada pela taxa de oscilação do próprio oscilador silábico, que apresenta todas as propriedades rítmicas da fala natural.



Scanned with CamScanner

## 6.4 O papel da prosódia em modelos de percepção da fala

Os modelos computacionais de *reconhecimento da fala* se compõem normalmente de um modelo acústico e de um modelo de língua. O primeiro aprende a associar determinadas características acústicas gerais com os fonemas da língua, sendo que esses modelos podem incorporar ou não parâmetros fonéticos clássicos como formantes, F0, duração, entre outros. Quando não incorporam diretamente o conhecimento fonético-linguístico, utilizam técnicas de processamento de sinal da fala que extraem informação acústica geral como no caso de coeficientes mel-cepstrais a cada 20 ms, por exemplo. Embora baseados em propriedades do sistema auditivo, esses coeficientes são dificilmente associados a parâmetros fonético-linguísticos como formantes. O modelo de língua restringe as sequências de fonemas possíveis em função da *fonotaxe* e mesmo da sintaxe da língua. A fonotaxe rege as sequências de sons permitidas na língua em questão (por exemplo, o português não tem sequência de /s/ seguida de consoante em início de sílaba).

Os modelos de reconhecimento de fala, no entanto, em geral limitados em termos de vocabulário a ser reconhecido ou requerendo uma fase de aprendizado para cada nova pessoa que deseja usar o sistema, não têm uma preocupação em simular o processo de percepção da fala e, portanto, ignoram em sua maioria o papel da prosódia para guiar essa percepção.

Ao leitor atento fica evidente do exposto até aqui que as duas funções básicas da prosódia, a de segmentação e a de proeminência, têm um papel importante para guiar a percepção da fala. A função de proeminência chama a atenção do ouvinte para um trecho de fala, normalmente uma palavra, o que guia a recuperação daquilo que é dito, do que se fala, de um tema, por exemplo. Num diálogo, quando um dos participantes chama a atenção para determinado elemento de sua fala realizando uma ênfase, servindo-se de foco estreito, ele guia o interlocutor na decodificação de sua intenção comunicativa.

A segmentação da fala realizada pela prosódia, por outro lado, facilita a interpretação daquilo que se diz, não apenas nos aspectos para- e extralinguísticos, como atitudes e afetos, mas do próprio conteúdo informacional, como vimos na seção sobre prosódia e sintaxe. Uma pausa e um movimento melódico distinto entre os trechos "não" e "fiz isso" são suficientes para mudar aquilo que era a negação de um ato ("Não fiz isso.") para a confirmação de uma suposição equivocada sobre a não realização desse mesmo ato em que o "não" que antecede a pausa nega o que foi afirmado pelo interlocutor.

As teorias de percepção da fala consideram uma comparação entre a informação acústica presente na fala, da qual se extraem pistas que, aliadas à informação da língua de que o ouvinte dispõe, permite a decodificação da fala. O tipo de processamento que nosso encéfalo faz com o sinal de fala que ouvimos começa pela extração das características do som no sistema periférico auditivo e sobe até as vias auditivas superiores, processamento que se chama de *bottom-up*. Mas, para entendermos o que e como foi dito, nosso encéfalo utiliza a informação do léxico mental e de elementos sonoros armazenados no nosso substrato neural para fazer hipóteses que guiam a interpretação do sinal de fala, processamento esse que se chama de *top-down*. A maior parte das teorias consideram que as pistas extraídas do sinal de fala são de natureza acústica ou fonético-acústica. No entanto, a **teoria motora de percepção da fala** considera que as pistas extraídas são de natureza articulatória, isto é, recuperamos os movimentos dos articuladores que tornaram a variação acústica possível. Nesse tipo de teoria, a integração entre elementos sonoros e visuais da enunciação como os movimentos orofaciais são mais simples, uma vez que os dois trabalhariam no mesmo espaço de coordenadas: movimento de articulador. Nenhuma das teorias integra de forma transparente a prosódia da fala, a não ser de forma geral nas linhas do que falamos nos parágrafos precedentes.

Teorias generalistas ou holísticas consideram a percepção da prosódia como sendo uma espécie de forma ou molde geral para a decodificação de afetos, atitudes e outros aspectos expressivos. Nessas teorias, não somente a melodia realizada por padrões de F0 tem um papel importante na interpretação dos enunciados, mas também a qualidade de voz e o ritmo da fala. São advogados dessa visão geral da prosódia percebida o americano Bolinger e o húngaro Fónagy.



Scanned with CamScanner

O objetivo da concepção de teorias e da construção de modelos de produção e percepção da prosódia é o de fornecer uma maneira de pensar a entoação e o ritmo da fala e os componentes associados a uma e outra como princípios organizadores da fala, como vimos nos capítulos 3 e 4. É também uma maneira de estabelecer hipóteses sobre o funcionamento de nosso substrato neural ao processar a informação prosódica, bem como estabelecer hipóteses sobre as relações entre a prosódia e os elementos linguísticos, discursivos e expressivos, como vimos no capítulo 5. As hipóteses novas nos permitem pensar fatos novos não apenas do ponto de vista linguístico, como também do ponto de vista do uso linguístico por parte de um indivíduo, permitindo-nos pensar em aplicações da prosódia para a área forense bem como para a área de ensino de língua estrangeira.

CAPÍTULO 7

# Para saber mais

Antes de recomendar leituras importantes sobre a prosódia da fala, é importante apontar listas de discussão de grupos de pesquisa na área, mesmo porque circula nessas discussões questões relativas à literatura.

Fundado em 2004, o grupo Yahoo ProFala foi formado a partir da iniciativa de quatro grupos de pesquisa na área de prosódia, o LABFON, da Faculdade de Letras da UFMG, o Laboratório de Fonética Acústica, da Faculdade de Letras da UFRJ, o LIACC, da PUC-SP, e o Grupo de Estudos de Prosódia da Fala, do Instituto de Estudos da Linguagem da Unicamp. Ele tem atualmente 161 membros e discute, em português, assuntos relacionados à prosódia da fala e divulga eventos relacionados ao tema. É preciso se inscrever para entrar no grupo em: <https://br.groups.yahoo.com/neo/groups/profala/info>.

O grupo SProSIG, associado à International Speech Communication Association, visa à promoção, ao estudo e à difusão do conhecimento da prosódia em geral, com informação e interação em inglês. O *site* do SProSIG se encontra aqui: <http://sprosig.isle.illinois.edu/>. Há um grupo de discussão, também no Yahoo, que pode ser acessado aqui: <https://uk.groups.yahoo.com/neo/groups/speech_prosody/info>.

O SProSIG organiza, em anos pares, a conferência internacional Speech Prosody, cuja série começou em 2002 em Aix-en-Provence, tendo sido realizada em Campinas em 2008. O mesmo grupo mantém à disposição os anais de todas as conferências, fonte única para



Scanned with CamScanner

acompanhar os temas de pesquisa recorrentes sobre prosódia da fala em âmbito mundial.

No Brasil, desde 2010, organizaram-se em anos pares, até o momento, as Escolas de Prosódia, com o objetivo de formar pesquisadores para a área a partir de cursos oferecidos por profissionais. Em anos ímpares, por outro lado, os Colóquios de Prosódia da Fala possibilitam aos pesquisadores da área apresentarem seus trabalhos e pesquisadores de diversas gerações se encontrarem. Ambos os eventos têm sido promovidos pela Associação Luso-Brasileira de Ciências da Fala (LBASS), cujo *site* é: <http://www.lbass.org/>.

A LBASS disponibiliza em seu *site* os anais do VI Colóquio Brasileiro de Prosódia da Fala, com artigos que cobrem os temas apresentados neste livro, bem como interfaces como prosódia da leitura, prosódia e música, prosódia e estilo telefônico, a prosódia do sarcasmo e de ironia, qualidade de voz, prosódia e atitudes, percepção da prosódia, variação prosódica, prosódia visual, marcadores discursivos em espanhol, aquisição da prosódia, além de outros. Para um aprofundamento nos temas discutidos neste livro, recomendamos ainda o que segue.

## 7.1 Capítulo 1

Para o estudo da prosódia da fala em geral, não há muitos livros-texto, sendo a maior parte coletâneas de artigos.

Em português, temos a coletânea *Estudos de prosódia*, organizada por Scarpa e que aborda em seus artigos as interfaces da prosódia da fala com a psicolinguística, a fonologia, a aquisição da linguagem, a fonoaudiologia, além de trabalhos sobre análise e síntese da prosódia, de fonética acústica e articulatória. Além desse, o recente livro *Prosódia da fala*, organizado por Freitag e Lucente, apresenta temas novos e de ponta, abordando a prosódia de língua estrangeira, a pesquisa da prosódia na área de fonoaudiologia, os aspectos entoacionais da prosódia, prosódia e variação, ensino de prosódia e prosódia articulatória, campo inovador, com poucos trabalhos. Em seu capítulo 5, o livro de Barbosa e Madureira (2015) aprofunda os aspectos metodológico-experimentais em torno da prosódia, além de apresentar o conceito de formante visto aqui e as técnicas de análise acústica em sua primeira parte, que inclui como fazer a segmentação dos sons da fala.

Em inglês, o livro *Suprasegmentals,* de Lehiste (1970), é fundamental para quem queira ter uma boa formação na área, porque apresenta os princípios metodológicos dos quinze anos que o precederam, as primeiras questões da área, as primeiras comparações entre línguas. Trata dos parâmetros prosódicos do ponto de vista da produção e da percepção, estabelecendo a conexão com sua função linguística, discutindo nesse contexto especialmente a questão do acento (*stress*). Entre outros trabalhos seminais sobre prosódia, recomendamos os seguintes autores (e obras) que foram fundamentais para dar base às gerações de prosodistas que os sucederam: Beckman (1996), Bolinger (1958, 1978, 1986, 1989), Campbell (1992), Crystal (1969), Cutler (1982, 1983, 1994, 1996), Fant e Kruckenberg (1996), Fónagy (1983), Fraisse (1963), Fry (1955, 1958), Fujisaki e Hirose (1984), Fujimura (1992), Gibbon (1976), Goldman-Eisler (1967, 1972), Grosjean (1983), Hirschberg (2002), Hirst (2006), Nooteboom (1997), Ohala (1977), Pierrehumbert (1980), Pike (1945), Rapp (1971), Rossi (2000), Vaissière (1983), sem contar os pesquisadores em seus respectivos países, notadamente na Europa, nos Estados Unidos e no Japão.

Recomendamos ainda as coletâneas organizadas por Oliver Niebuhr, *Understanding Prosody*, que apresenta pesquisa atual em algumas línguas, aplicação de modelos da entoação, variação prosódica regional, prosódia da conversa, diferenças prosódicas interlinguísticas, entre outros assuntos. Outra coletânea em inglês, bem anterior a essa, *Prosody: Models and Measurements*, foi organizada por Cutler e Ladd, explora aspectos como a universalidade potencial de alguns parâmetros prosódicos, modelo de produção da entoação, prosódia da estrutura informacional, além de outros. Aspectos metodológicos da pesquisa experimental em prosódia da fala são apresentados na coletânea organizada por editores coordenados por Sudhoff, *Methods in Empirical Prosody Research*. Nela se discutem de forma mais aprofundada vários temas tocados em nosso livro, como delimitação de duração, estilização de curvas de F0, qualidade de voz, ritmo da



Scanned with CamScanner

fala, tarefas para eliciar *corpora* e métodos para estudos de prosódia em seus aspectos de produção e de percepção.

Em francês, recomendamos os excelentes livros de Lacheret-Dujour e Beaugendre, *La Prosodie du français*, e de Di Cristo, *La Prosodie de la parole*. Ambos apresentam dos conceitos básicos aos mais avançados, cobrindo e aprofundando parte dos conceitos vistos neste livro introdutório: o primeiro discorre largamente sobre modelos prosódicos e aspectos metodológicos. O segundo apresenta uma larga introdução, bem como cobre extensamente os aspectos fonéticos e linguísticos da prosódia da fala com interesse também por modelos da prosódia e sua relação com o significado.

O livro *Foundations of Voice Studies*, de Kreiman e Siditis (2011), é uma excelente referência para o aprofundamento do estudo sobre qualidade de voz. Os demais correlatos físicos e perceptivos da prosódia encontram um aprofundamento importante no livro *Acoustic and Auditory Phonetics*, de Johnson (2012) já na 3ª edição.

Sobre o caráter composicional da percepção da prosódia, foi Cumming (2011) que confirmou um resultado anteriormente conhecido (Rosen, 1977), o de que os ouvintes ouvem estímulos em que o valor de F0 varia como sendo mais longos do que aqueles em que a F0 é constante. E mostrou que o *pitch* de tons agudos aumenta com a intensidade, isto é, a sensação de agudo é tanto maior em tons agudos quanto maior a intensidade e, mais tarde, que à medida que a duração diminui, a sensação de grave aumenta para o mesmo valor de F0 (Stevens, 1939).

*Corpora* de fala espontânea para análise prosódica, em trabalhos que privilegiam esse componente da fala como organizador dos atos de fala, são disponibilizados por Cresti e Moneglia (2005) e Raso e Mello (2012).

Excelentes introduções sobre forma e função prosódicas são os artigos de Hirst (2005) e o de Xu (2005), que ainda faz essa relação com a articulação. Em relação à importância da experimentação em prosódia da fala, ler o artigo de Xu (2010) em defesa da fala de laboratório.

## 7.2 Capítulo 2

O artigo de Botinis e colaboradores (2001) dá um excelente panorama sobre as unidades prosódicas e sua sinonímia, bem como sobre as últimas pesquisas na área até a data. Fant (1991) discorre amplamente sobre as unidades de sílaba, palavra e grupo acentual com base no sueco.

Especificamente sobre a sílaba como unidade básica da prosódia relacionada à transição C-V temos o livro *The PIVOT Model of Speech Parsing*, de Dogil e Braun (1988), e o livro *Incursões em torno do ritmo da fala*, de Barbosa (2006). Esse mesmo autor recentemente resenhou o trabalho seminal de Stetson (1951) a respeito da preponderância da sílaba (Barbosa, 2017).

Para aprofundar o conhecimento sobre a sílaba nas línguas românicas, recomendamos a leitura do capítulo de Heinz (2014) e, para entender melhor a noção de mora, recomendamos o artigo de Kubozono (1989).

## 7.3 Capítulo 3

O livro de Barbosa (2006), *Incursões em torno do ritmo da fala*, aprofunda o tema do ritmo do ponto de vista prosódico a partir de um modelo de osciladores acoplados. A tese que defende, de uma dinâmica oscilatória subjazendo a esse ritmo, na contracorrente de abordagens que levam em conta apenas a duração segmental, é partilhada por trabalhos como os de Cummins e Port (1998) e, mais recentemente, por trabalhos de pesquisadoras bem ativas na área de prosódia como Wagner e Malisz e seus colaboradores (Windmann, Šimko e Wagner, 2015; Inden *et al.* 2012).

Recomenda-se a leitura de textos de orientação de metodologia experimental que discorrem sobre como realizar procedimentos de aleatorização dos estímulos, o uso de estímulos distratores que dificultem ao ouvinte o uso de estratégias enviesadas para dar sua resposta, número de estímulos para que o ouvinte não canse, além de procedimentos estatísticos para avaliar a coerência de respostas



Scanned with CamScanner

fala, tarefas para eliciar *corpora* e métodos para estudos de prosódia em seus aspectos de produção e de percepção.

Em francês, recomendamos os excelentes livros de Lacheret-Dujour e Beaugendre, *La Prosodie du français*, e de Di Cristo, *La Prosodie de la parole*. Ambos apresentam dos conceitos básicos aos mais avançados, cobrindo e aprofundando parte dos conceitos vistos neste livro introdutório: o primeiro discorre largamente sobre modelos prosódicos e aspectos metodológicos. O segundo apresenta uma larga introdução, bem como cobre extensamente os aspectos fonéticos e linguísticos da prosódia da fala com interesse também por modelos da prosódia e sua relação com o significado.

O livro *Foundations of Voice Studies*, de Kreiman e Siditis (2011), é uma excelente referência para o aprofundamento do estudo sobre qualidade de voz. Os demais correlatos físicos e perceptivos da prosódia encontram um aprofundamento importante no livro *Acoustic and Auditory Phonetics*, de Johnson (2012) já na 3ª edição.

Sobre o caráter composicional da percepção da prosódia, foi Cumming (2011) que confirmou um resultado anteriormente conhecido (Rosen, 1977), o de que os ouvintes ouvem estímulos em que o valor de F0 varia como sendo mais longos do que aqueles em que a F0 é constante. E mostrou que o *pitch* de tons agudos aumenta com a intensidade, isto é, a sensação de agudo é tanto maior em tons agudos quanto maior a intensidade e, mais tarde, que à medida que a duração diminui, a sensação de grave aumenta para o mesmo valor de F0 (Stevens, 1939).

*Corpora* de fala espontânea para análise prosódica, em trabalhos que privilegiam esse componente da fala como organizador dos atos de fala, são disponibilizados por Cresti e Moneglia (2005) e Raso e Mello (2012).

Excelentes introduções sobre forma e função prosódicas são os artigos de Hirst (2005) e o de Xu (2005), que ainda faz essa relação com a articulação. Em relação à importância da experimentação em prosódia da fala, ler o artigo de Xu (2010) em defesa da fala de laboratório.

## 7.2 Capítulo 2

O artigo de Botinis e colaboradores (2001) dá um excelente panorama sobre as unidades prosódicas e sua sinonímia, bem como sobre as últimas pesquisas na área até a data. Fant (1991) discorre amplamente sobre as unidades de sílaba, palavra e grupo acentual com base no sueco.

Especificamente sobre a sílaba como unidade básica da prosódia relacionada à transição C-V temos o livro *The PIVOT Model of Speech Parsing*, de Dogil e Braun (1988), e o livro *Incursões em torno do ritmo da fala*, de Barbosa (2006). Esse mesmo autor recentemente resenhou o trabalho seminal de Stetson (1951) a respeito da preponderância da sílaba (Barbosa, 2017).

Para aprofundar o conhecimento sobre a sílaba nas línguas românicas, recomendamos a leitura do capítulo de Heinz (2014) e, para entender melhor a noção de mora, recomendamos o artigo de Kubozono (1989).

## 7.3 Capítulo 3

O livro de Barbosa (2006), *Incursões em torno do ritmo da fala*, aprofunda o tema do ritmo do ponto de vista prosódico a partir de um modelo de osciladores acoplados. A tese que defende, de uma dinâmica oscilatória subjazendo a esse ritmo, na contracorrente de abordagens que levam em conta apenas a duração segmental, é partilhada por trabalhos como os de Cummins e Port (1998) e, mais recentemente, por trabalhos de pesquisadoras bem ativas na área de prosódia como Wagner e Malisz e seus colaboradores (Windmann, Šimko e Wagner, 2015; Inden *et al.* 2012).

Recomenda-se a leitura de textos de orientação de metodologia experimental que discorrem sobre como realizar procedimentos de aleatorização dos estímulos, o uso de estímulos distratores que dificultem ao ouvinte o uso de estratégias enviesadas para dar sua resposta, número de estímulos para que o ouvinte não canse, além de procedimentos estatísticos para avaliar a coerência de respostas



Scanned with CamScanner

do mesmo ouvinte ou entre ouvintes, bem como em levar em conta a possibilidade de uso excessivo de falsos alarmes, para acertar respondendo da mesma maneira. A esse respeito, sugerem-se as leituras de McGuire (2010) e Goldman-Eisler (1968). Para métodos experimentais específicos para a avaliação de proeminência e segmentação prosódicas, recomendamos os trabalhos de Barbosa (2006, 2008, 2010) e de Cole *et al.* (2008).

Sobre o conceito e os limites da memória de trabalho, consultar o livro de Cowan (1997).

Sobre ênfase, ver os trabalhos de Kohler (2006) e Niebuhr (2010).

Um programa disponível para realizar a delexicalização bem usado na literatura é o PURR (Sonntag e Portele, 1998).

Quanto à noção de ritmo dentro de um quadro filosófico, remetemos o leitor a Whitehead (1919) e a Sauvanet (2000).

Sobre a produção do ritmo da fala, os artigos seguintes apresentam resultados experimentais importantes na literatura: Allen (1975), Beckman (1992), Bell-Berti e Harris (1991), Eriksson (1991), De Jong *et al.* (1993) e Edwards *et al.* (1991). O trabalho de O'Shaughnessy (1985) estuda o ritmo na fala espontânea e o de Grosjean e Collins (1979), a relação entre ritmo da fala e respiração.

Sobre segmentação da fala com base em princípios rítmicos e linguísticos, recomendamos a leitura de Farnetani (1989) para o italiano e Cutler (1991) para o inglês.

Sobre a dicotomia entre línguas de ritmo acentual e silábico, ver os trabalhos de Lehiste (1977), Dauer (1983), Roach (1982), Kohler (1983), Nooteboom (1991), Wenk (1985), Barbosa (2006) e o controverso, porém importante artigo de Ramus *et al.* (1999).

Sobre a duração para a síntese da fala, ver o trabalho de Klatt (1987), e sobre síntese da entoação, ver Pierrehumbert (1981).

Sobre correlatos acústicos do acento, ver os trabalhos de Morton e Jassem (1965), Fry (1955, 1958) e Barbosa *et al.* (2013). As consequências acústicas da fronteira prosódica para o inglês foram investigadas primeiramente por Wightman *et al.* (1992).

As unidades V-V se relacionam com a noção de *perceptual-center*, um ponto de ancoragem da percepção das sílabas em torno da transição C-V, cujo trabalho seminal é o de Morton *et al.* (1976).

Sobre a percepção da prosódia, ver Shattuck-Hufnagel e Turk (1996) e Cutler *et al.* (1997). Sobre a percepção do ritmo da fala, ver Cutler (1994) e Vaissière (1991) e, especificamente sobre a percepção da taxa de elocução, ver os trabalhos de Miller e Grosjean (1981) e Rietveld e Gussenhoven (1987).

## 7.4 Capítulo 4

Para esse capítulo, recomendamos primeiramente aqueles que propõem sistemas de notação entoacional, como o INTSINT, de Hirst e Di Cristo (1998), o ToBI, de Silverman *et al.* (1992), e o DaTO, de Lucente (2012).

O artigo de Botinis e colaboradores (2001) cobre os principais aspectos da pesquisa sobre a entoação da fala, com várias referências bibliográficas.

O trabalho de t'Hart *et al.* (1990) apresenta um dos primeiros modelos de percepção e estilização da entoação.

O trabalho de Ladd (1984) dá um excelente panorama sobre a declinação de F0.

## 7.5 Capítulo 5

A teoria em que nos fundamentamos para segmentar os trechos da narrativa da lenda amazônica é a de Grosz e Sidner (1986), que procura identificar intenções e mecanismos atencionais que permitem a estruturação do discurso. Recomendamos ainda o trabalho de Wennerstrom (2001).

Os trabalhos mencionados na seção de prosódia e processamento são os de Baum (1998), Belin *et al.* (1998), Dapretto *et al.* (1999), Dogil *et al.* (2002), Mauk e Buonomano (2004), Ivry (1996), O'Boyle, Freeman e Cody (1996), Hickok e Poeppel (2000), Robin *et al.* (1990),



Scanned with CamScanner

Zatorre (1997), Shah *et al.* (2006), Nicholls (1996), Shapiro e Danly (1985) e Frühholz *et al.* (2011).

Sobre expressividade da fala, há uma grande lista de trabalhos relevantes, mas chamo a atenção para os trabalhos sobre o português brasileiro de Madureira (1996, 2004, 2011, 2016).

Sobre prosódia e discurso, ver os trabalhos de Hirschberg e Swerts listados nas referências.

## 7.6 Capítulo 6

Além dos modelos e teorias mencionados e listados nas referências, como os de MacNeilage (1998) e Shattuck-Hufnagel e Klatt (1979), recomendamos ainda o livro de Levelt (1989) e a discussão de Lindblom e MacNeilage (1986) sobre a teoria da ação.

# Referências bibliográficas

ALLEN, G. D. (1975). Speech Rhythm: Its Relation to Performance Universals and Articulatory Timing. *Journal of Phonetics*, 3, 75-86.

ARANTES, P. (2010). *Integrando produção e percepção de proeminências secundárias numa abordagem dinâmica do ritmo da fala*. Tese de doutorado. Campinas: Universidade Estadual de Campinas.

AZIZ-ZADEH, L., SHENG, T., & GHEYTANCHI, A. (2010). Common Premotor Regions for the Perception and Production of Prosody and Correlations with Empathy and Prosodic Ability. *PloS one*, *5*(1), e8759.

BARBOSA, P. A. (2006). *Incursões em torno do ritmo da fala*. Campinas: Pontes.

___ (2008). Prominence- and Boundary-related Acoustic Correlations in Brazilian Portuguese Read and Spontaneous Speech. *Proc. of the Speech Prosody 2008 Conference*. Campinas: RG, 257-260.

___ (2010). Automatic Duration-related Salience Detection in Brazilian Portuguese Read and Spontaneous Speech. *Proc. Speech Prosody 2010*, Chicago (100067:1-4).

___ (2017). RH Stetson, Motor Phonetics: A Study of Speech Movements in Action, Amsterdam, North Holland Publishing Co., 1951. *Phonetica*, 74(4), 255-258.

BARBOSA, P. A.; ERIKSSON, A.; ÅKESSON, J. (2013). Cross-linguistic Similarities and Differences of Lexical Stress Realisation in Swedish and Brazilian Portuguese. In ASU, E. L.; LIPPUS, P. (*orgs.*). *Nordic Prosody. Proceedings from the XIth Conference*, Tartu 2012 (pp. 97-106). Frankfurt am Main: Peter Lang,

BARBOSA, P. A.; MADUREIRA, S. (2015). *Manual de fonética acústica experimental: aplicações a dados do português*. São Paulo: Cortez.

BAUM, S. R. (1998). The Role of Fundamental Frequency and Duration in the Perception of Linguistic Stress by Individuals with Brain Damage. *Journal of Speech, Language, and Hearing Research*, *41*(1), 31-40.

BECKMAN, M. E. (1992). Evidence for Speech Rhythms across Languages. IN TOHKURA, Y.; VATIKIOTIS-BATESON, E.; SAGISAKA, Y. (*orgs.*). *Speech Perception, Production and Linguistic Structure*. Amsterdam: IOS Press, 457-463.

___ (1996). The Parsing of Prosody. *Language and Cognitive Processes*, 11(1-2), 17-68.

BELIN, P. *et al.* (1998). Lateralization of Speech and Auditory Temporal Processing. *J. Cogn. Neurosci.* 10, 536–540.

BELL-BERTI, F.; HARRIS, K. (1981). A Temporal Model of Speech Production. *Phonetica*, 38, 9-20.



Scanned with CamScanner

BOLINGER, D. L. (1958). A Theory of Pitch Accent in English. *Word*, 14(2-3), 109-149.
___ (1978). Intonation Across Languages. Universals of Human. *Language*, 2, 471-524.
___ (1986). *Intonation and its Parts: Melody in Spoken English*. Stanford: Stanford University Press.
___ (1989). *Intonation and its Uses: Melody in Grammar and Discourse*. Stanford: Stanford University Press.
BOERSMA, P.; WEENINK, D. (2018). Praat: Doing Phonetics by Computer (versão 6.0.42) [Programa de computador].
BOTINIS, A., GRANSTRÖM, B., MÖBIUS, B. (2001). Developments and Paradigms in Intonation Research. *Speech Communication*, 33, p. 263-296.
BYRD, D.; SALTZMAN, E. (2003). The Elastic Phrase: Modeling the Dynamics of Boundary-adjacent Lengthening. *Journal of Phonetics*, 31(2), 149-180.
CAMPBELL, W. N. (1992). Syllable-based Segment Duration. In: BAILLY, G., BENOIT, C., SAWALIS, T. R. (eds). *Talking Machines*. Amsterdam: North-Holland, p. 211-224.
COLE, J.; GOLDSTEIN, L.; KATSIKA, A.; MO, Y.; NAVA, E.; TIEDE, M. (2008). Perceived Prosody: Phonetic Bases of Prominence and Boundaries. *J. Acoust. Soc. Am.* 24(4), 2496.
CRESTI, E. & MONEGLIA, M. (*orgs.*) (2005). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: John Benjamins.
CRYSTAL, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.
CUMMING, R. (2011). The Effect of Dynamic Fundamental Frequency on the Perception of Duration. *Journal of Phonetics*, 39(3): 375–387.
CUMMINS, F.; PORT, R. F. (1998). Rhythmic Constraints on Stress Timing in English. *Journal of Phonetics*, 26(2): 145-171.
CUTLER, A. (org.). (1982). *Slips of the Tongue and Language Production*. Haia: Mouton.
CUTLER, A. (1983). Speakers' Conceptions of the Functions of Prosody. IN CUTLER, A.; LADD, D. R. (*Orgs.*). *Prosody: Models and Measurements*. Heidelberg: Springer, pp. 79-91.
CUTLER, A. (1991). Linguistic Rhythm and Speech Segmentation. In: SUNDEBERG, J., NORD, L. & CARLSON, R. (eds). *Music, Language, Speech and Brain*. Macmillan: Houndsmills & London, p. 157-166.
___ (1994). The Perception of Rhythm in Language. *Cognition*, 50, 79-81.
___ (1996). Prosody and the Word Boundary Problem. In MORGAN, J. L.; DEMUTH, K. (orgs.). *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. Hillsdale: Lawrence Erlbaum Associates, p. 87-99.
CUTLER, A.; DAHAN, D.; VAN DONSELAAR, W. (1997). Prosody in the Comprehension of Spoken Language: A Literature Review. *Language and Speech*, 40(2), 141-201.
CUTLER, A.; ISARD, S. D. (1980). The Production of Prosody. In BUTTERWORTH, B. (org.). *Language Production: Speech and Talk*. Londres, Nova York, Toronto, Sydney & San Francisco: Academic Press, p. 245-269.
DAPRETTO, M.; HARIRI, A.; BIALIK, M.; BOOKHEIMER, S. (1999). Cortical Correlates of Affective vs. Linguistic Prosody: an fMRI Study. *NeuroImage*, 9(6), S1054.
DAUER, R. (1983). Stress-timing and Syllable Timing Revisited. *Journal of Phonetics*, 11, 51-62.
DE JONG, K., BECKMAN, M. E.; EDWARDS, J. (1993). The Interplay between Prosodic Structure and Coarticulation. *Language and Speech*, 36(2-3), 197-212.
DOGIL, G., BRAUN, G. (1988). *The PIVOT Model of Speech Parsing*. Viena: Austrian Academy of Sciences Press.
DOGIL, G.; ACKERMANN, H.; GRODD, W.; HAIDER, H.; KAMP, H.; MAYER, J.; RIECKER, A.; WILDGRUBER, D. (2002). The Speaking Brain: a Tutorial Introduction to fMRI Experiments in the Production of Speech, Prosody and Syntax. *Journal of Neurolinguistics* 15, 59-90.
EDWARDS, J.; BECKMAN, M. E.; FLETCHER, J. (1991). The Articulatory Kinematics of Final Lengthening. *J. Acoust. Soc. Am.* 89 (1), 369-382.
FANT, G. (1991). Units of Temporal Organization. Stress Groups versus Syllables and Words. *Proceedings of the XIIth ICPhS*, Aix-en-Provence, p. 247-250.
FANT, G.; KRUCKENBERG, A.; NORD, L. (1991). Durational Correlates of Stress in Swedish, French and English. *Journal of Phonetics*, 19 (3/4), 351-365.
___. (1991). Prosodic and Segmental Speaker Variations. *Speech Communication*, 10(5-6), 521-531.
FANT, G.; KRUCKENBERG, A. (1996). On the Quantal Nature of Speech Timing. In *Proceedings of the Fourth International Conference on Spoken Language*, vol. 4, p. 2044-2047.
FARNETANI, E. (1989). Acoustic Correlates of Linguistic Boundaries in Italian: a Study on Duration and Fundamental Frequency. *Proc. of Eurospeech*, 2, p. 332-335.
FÓNAGY, I. (1983). *La vive voix: essais de psycho-phonétique*. Paris: Payot.
FÓNAGY, I.; FÓNAGY, J. (1983). L'intonation et l'organisation du discours. *Bulletin de la Société de Linguistique de Paris*, 78, 161-209.
FRAISSE, P. (1963). *The Psychology of Time*. Nova York: Harper and Row.
FRÜHHOLZ, S., CERAVOLO, L., & GRANDJEAN, D. (2011). Specific Brain Networks During Explicit and Implicit Decoding of Emotional Prosody. *Cerebral Cortex*, *22*(5), 1107-1117.
FRY, D. B. (1955). Duration and Intensity as Physical Correlates of Linguistic Stress. *JASA*, 27, 765-768.
___ (1958). Experiments in the Perception of Stress. *Language and Speech*, 1, 120-152.
FUJIMURA, O. (1992). Phonology and Phonetics. A Syllable-based Model of Articulatory Organization. *Journal of the Acoustical Society of Japan* (E), 13(1), 39-48.
FUJISAKI, H.; HIROSE, K. (1984). Analysis of Voice Fundamental Frequency Contours for Declarative Sentences of Japanese. *Journal of the Acoustical Society of Japan* (E), 5(4): 233-241.
GIBBON, D. (1976). *Perspectives of Intonation Analysis*. Berna: Lang.
GOLDMAN-EISLER, F. (1967). Sequential Temporal Patterns and Cognitive Processes in Speech. *Quarterly Journal of Experimental Psychology*, 10, 96-106.
___ (1968). *Psycholinguistics: Experiments in Spontaneous Speech*. Nova York: Academic Press.
___1972). Pauses, Clauses, Sentences. *Language and Speech*, 15, 103- 113.
GROSJEAN, F. (1983). How Long is the Sentence? Prediction and Prosody in the On-line Processing of Language. *Linguistics* 21 (3), 501, 530.
GROSJEAN, F.; COLLINS, M. (1979). Breathing, Pausing and Reading. *Phonetica*, 36, 98-114.
GROSJEAN, F.; GROSJEAN, L.; LANE, H. (1979). The Patterns of Silence: Performance Structures in Sentence Production. *Cognitive Psychology*, 11(1), 58-81.
GROSZ, B. J., SIDNER, C. L. (1986). Attention, Intention, and the Structure of Discourse. *Comp. Ling.*, 12: 175–204.
HEINZ, M. (2014). Syllable Complexity in the Diachrony of Romance Languages: A Center vs. Periphery View and the Syllable vs. Word Rhythm Paradigm. IN CARO REINA, J., SZCZEPANIAK, R. (orgs.). *Syllable and Word Languages*. Berlin, Boston: De Gruyter, p. 87-111.
HICKOK, G.; POEPPEL, D. (2000). Towards a Functional Neuroanatomy of Speech Perception. *Trends in Cognitive Sciences*, 4 (4), 131-138.



Scanned with CamScanner

HIROSE, K.; FUJISAKI, H. (1982). Analysis and Synthesis of Voice Fundamental Frequency Contours of Spoken Sentences. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'82* (vol. 7, pp. 950-953). IEEE.

HIRSCHBERG, J. (2002). Communication and Prosody: Functional Aspects of Prosody. *Speech Communication*, 36(1-2), 31-43.

HIRST, D. (2005). Form and Function in the Representation of Speech Prosody In: Hirose, K.; HIRST, D.; SAGISAKA, Y. (eds). *Quantitative Prosody Modeling for Natural Speech Description and Generation. Speech Communication*, 46 (3-4), 334-347.

___ (2006). Phonetics. The Prosody of Speech and Language. In BROWN, K. (*org.*). *Encyclopaedia of Language and Linguistics*. Oxford: Oxford University Press, vol. 10, p. 167-178.

HIRST, D.; DI CRISTO, A. (orgs.) (1998). *Intonation Systems: a Survey of Twenty Languages*. Cambridge: Cambridge University Press.

INDEN, B.; MALISZ, Z.; WAGNER, P.; WACHSMUTH, I. (2012). Rapid Entrainment to Spontaneous Speech: A Comparison of Oscillator Models. In *Proc. of the Annual Meeting of the Cognitive Science Society*, 34 (34).

KLATT, D. H. (1987). Review of Text-to-speech Conversion for English. *The Journal of the Acoustical Society of America*, *82*(3), 737-793.

KOHLER, K. J. (1983) Stress-timing and Speech Rate in German. A Production Model. *Arbeitsberichte des Instituts fur Phonetik des Universitat Kiel*, pp. 7-53.

___ (2006). What is Emphasis and How is it Coded? *Proc. of the 2006 Speech Prosody Conference*, Dresden, 748-751.

KOHNO, M. (1992). Two Mechanisms of Processing Sound Sequences. In: TOHKURA, Y.; VATIKIOTIS-BATESON, E.; SAGISAKA, Y. (orgs.). *Speech, Perception, Production and Linguistic Structure*. Amsterdam: IOS Press, p. 287-293.

KREIMAN, J.; SIDTIS, D. (2011). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Malden: John Wiley & Sons.

KUBOZONO, H. (1989). The Mora and Syllable Structure in Japanese: Evidence from Speech Errors. *Language and Speech*, *32*(3), 249-278.

LADD. D. R. (1984). Declination: a Review and Some Hypotheses. *Phonology Yearbook*, 1, 53-74.

LADD. D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

LEHISTE, I. (1970). *Suprasegmentals*. Cambridge, Mass.: MIT Press.

___ (1977). Isochrony Reconsidered. *Journal of Phonetics*, 5, 253-263.

LEVELT, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, Mass.: MIT Press.

LINDBLOM, B. E. F.; MACNEILAGE, P. (1986). Action Theory: Problems and Alternative Approaches. *Journal of Phonetics*, 14, 117-132.

LUCENTE, L. (2012). *Aspectos dinâmicos da fala e da entoação no português brasileiro*. Tese (Doutorado em Linguística). Campinas: Universidade Estadual de Campinas.

MACNEILAGE, P. F. (1998). The Frame/Content Theory of Evolution of Speech Production. *Behavioral and Brain Sciences*, 21(4), 499-511.

MADUREIRA, S. (1991). A qualidade de voz como recurso fonoestilístico. *Intercâmbio* (Porto), São Paulo, v. 1, p. 99-120.

___ (1996). A materialidade fônica, os efeitos de sentido e os papéis do falante. *D.E.L.T.A.*, São Paulo, v. 12, n.1, p. 87-93.

___ (2004). A expressão de atitudes e emoções na fala. In: KIRILLOS, L. (org.). *Expressividade*. São Paulo: Revinter, p. 15-25.

___ (2011). The Investigation of Speech Expressivity. In: MELLO, H.; PANUNZI, A.; RASO, T. (eds). *Illocution, Modality, Attitude, Information Patterning and Speech Annotation*. Firenze: Firenze University Press, v. 1, p. 101-118.

MADUREIRA, S. (2016). Intonation and Variation: the Multiplicity of Forms and Senses. *Dialectologia*, n. especial,v. VI, p. 54-74.

MARTIN, P. (2004). Winpitch-pro — A Tool for Text to Speech Alignment and Prosodic Analysis, *Proc. of the 2004 Speech Prosody Conference*, p. 545-548.

MCGUIRE, G. (2010). A Brief Primer on Experimental Designs for Speech Perception Research. Acessível em: <https://dokumen.tips/documents/a-brief-primer-on-experimental-designs-for-speech-gmcguir1experimentdesignspdfa.html>

MERTENS, P. (2004). Un outil pour la transcription de la prosodie dans les corpus oraux. *Traitement Automatique des langues* 45 (2), 109-130. *site*: https://sites.google.com/site/prosogram/home.

MILLER, J. L.; GROSJEAN, F. (1981). How the Components of Speaking Rate Influence Perception of Phonetic Segments. *J. Experimental Psychology*: Human Perception and Performance 7 (1), 208-215.

MORTON, J.; JASSEM, W. (1965). Acoustic Correlates of Stress. *Language and Speech*, 8, 159-181.

MORTON, J.; MARCUS, S.; FRANKISH, C. (1976). Theoretical Note: Perceptual Centers (P-Centers). *Psychological Review* 83 (5), 405-408.

NESPOR, M.; VOGEL, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.

NICHOLLS, M. E. R. (1996). Temporal Processing Asymmetries Between the Cerebral Hemispheres: Evidence and Implications. *Laterality* 1,97-137.

NIEBUHR, O. (2010). On the Phonetics of Intensifying Emphasis in German. *Phonetica*, 67(3), 170-198.

NOOTEBOOM, S. G. (1991). Some Observations on the Temporal Organisation and Rhythm of Speech. *Proceedings of the XIIth ICPhS*, Aix-en-Provence, 228-237.

___ (1997). The Prosody of Speech: Melody and Rhythm. In HARDCASTLE, W.; LAVER, J. (orgs.). *The Handbook of Phonetic Sciences*. Oxford: Blackwell, p. 640-673.

O'BOYLE, D. J.; FREEMAN, J. S.; CODY, F. W. (1996). The Accuracy and Precision of Timing of Self-paced, Repetitive Movements in Subjects with Parkinsons's Disease. *Brain*, v. 119, p. 51-70.

OHALA, J. J. (1977). The Physiology of Stress. In: HYMAN, L. M. (*Org.*). *Studies in Stress and Accent*. South California Occasional Papers in Linguistics, 4, 145-168.

O'SHAUGHNESSY, D. (1995). Timing Patterns in Fluent and Disfluent Spontaneous Speech. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, p. 600-603.

PIERREHUMBERT, J. B. (1980). *The Phonology and Phonetics of English Intonation*. Tese de doutorado. Cambridge, Mass.: MIT.

___ (1981). Synthesizing Intonation. *Journal of the Acoustical Society of America*, 70(4), 985-995.

PIKE, K. L. (1945). *The Intonation of American English: Linguistics*. Michigan: University of Michigan Press.

RAMUS, F.; NESPOR, M.; MEHLER, J. (1999). Correlates of Linguistic Rhythm in the Speech Signal. *Cognition*, 73, 265-292.

RAPP, K. (1971). A Study of Syllable Timing. *Speech Transmission Laboratory. Quaterly Progress and Status Report*, 1, 14-19.

RASO, T.; MELLO, H. (orgs.). (2012). *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte: UFMG.

RIETVELD, A. C. M.; GUSSENHOVEN, C. (1987). Perceived Speech Rate and Intonation. *J. Phonetics*, 15, 273-285.



Scanned with CamScanner

ROACH, P. (1982). On the Distinction Between 'Stress-timed' and 'Syllable-timed' Languages. In CRYSTAL, D. (org.). *Linguistic Controversies. Essays in Linguistic Theory and Practice*. Londres: Edward Arnold.

ROBIN, D. A. *et al.* (1990). Auditory Perception of Temporal and Spectral Events in Patients with Focal Left and Right Cerebral Lesions. *Brain Lang.* 39, 539-555.

ROSEN, S. M. (1977). The Effect of Fundamental Frequency Patterns on Perceived Duration. *Speech Transmission Laboratory — Quarterly Progress and Status Report, 18*, 17-30.

ROSSI, M. (2000). Intonation: Past, Present, Future. In *Intonation*. Dordrecht: Springer, p. 13-52.

SAUVANET, P. (2000). *Le Rythme et la raison*, vol. 1, *Rythmologiques*. Paris: Éditions Kimé.

SHAH, A. P.; BAUM, S. R.; DWIVEDI, V. D. (2006). Neural Substrates of Linguistic Prosody: Evidence from Syntactic Disambiguation in the Productions of Brain-Damaged Patients. *Brain and Language*, 96, 78–89.

SHAPIRO, B. E.; DANLY, M. The Role of the Right Hemisphere in the Control of Speech Prosody in Propositional and Affective Contexts. *Brain and Language* 25, nº 1 (1985): 19-36.

STEINHAUER, K.; ALTER, K.; FRIEDERICI, A. D. (1999). Brain Potentials Indicate Immediate Use of Prosodic Cues in Natural Speech Processing. *Nature Neuroscience* 2 (2), 191-196.

SHATTUCK-HUFNAGEL, S.; KLATT, D. (1979). The Limited Use of Distinctive Features and Markedness in Speech Production: Evidence from Speech Error Data. *Journal of Verbal Learning and Verbal Behavior*, v. 18, p. 41–55.

SHATTUCK-HUFNAGEL, S.; TURK, A. E. (1996). A Prosody Tutorial for Investigators of Auditory Sentence Processing. *Journal of Psycholinguistic Research* 25 (2), 193-247.

SILVERMAN, K.; BECKMAN, M.; PITRELLI, J.; OSTENDORF, M.; WIGHTMAN, C.; PRICE, P.; PIERREHUMBERT, J.; HIRSCHBERG, J. (1992). TOBI: A Standard for Labeling English Prosody. In *Proceedings of the International Conference on Speech and Language Processing* (ICSLP), v. 2, 867-870.

SONNTAG, G.; PORTELE, T. (1998). PURR – A Method for Prosody Evaluation and Investigation. *Computer Speech & Language*, 12(4), 437-451.

STETSON, R. H. (1951 [1928]). *Motor Phonetics: A Study of Speech Movements in Action*. Amsterdam: North Holland Publishing Co.

STEVENS, S. S. (1935). The Relation of Pitch to Intensity. *The Journal of the Acoustical Society of America* 6, 150. https://doi.org/10.1121/1.1915715

___ (1939). The Relation of Pitch to the Duration of a Tone . *The Journal of the Acoustical Society of America* 10, 255. https://doi.org/10.1121/1.1902092.

SWERTS, M. (1997). Prosodic Features at Discourse Boundaries of Different Strength. *The Journal of the Acoustical Society of America*101, 514-521.

SWERTS, M.; GELUYKENS, R. C. (1994) Prosody as a Marker of Intonation Flow in Spoken Discourse. *Language and Speech*, 37, 21-43.

T'HART, J.; COLLIER, R; COHEN, A. (1990). *A Perceptual Study of Intonation*. Cambridge: Cambridge University Press.

VAISSIÈRE, J. (1983). Language-independent Prosodic Features. In CUTLER, A.; LADD, R. (orgs.). *Prosody: Models and Measurements*. Berlin: Springer Verlag, p. 53-66.

___ (1991). Perceiving Rhythm in French? In *Proceedings of 12th International Congress of Phonetic Sciences*, 4, 258-261. Aix-en-Provence.

WENK, B. J. (1985). Speech Rhythms in Second Language Acquisition. *Language and Speech*, 28 (2), 157-175.

WENNERSTROM, A. (2001). *The Music of Everyday Speech: Prosody and Discourse Analysis*. Oxford: Oxford University Press.

WIGHTMAN, C. W.; SHATTUCK-HUFNAGEL, S.; OSTENDORF, M.; PRICE, P. (1992). Segmental Durations in the Vicinity of Prosodic Phrase Boundaries. *Journal of the Acoustical Society of America*, 91, 1707-1717.

WHITEHEAD, A. N. (1919). *An Enquiry concerning the Principles of Natural Knowledge*. Cambridge: Cambridge University Press.

WINDMANN, A.; ŠIMKO, J.; WAGNER, P. (2015). Optimization-based Modeling of Speech Timing. *Speech Communication* 74, 76-92.

XU, Y. (2005). Speech Melody as Articulatorily Implemented Communicative Functions. *Speech Communication*, 46(3-4), 220-251.

XU, Y. (2010). In Defense of Lab Speech. *Journal of Phonetics*, 38(3), 329-336.

ZWICKER, E.; Fastl, H. (1999). *Psychoacoustics: Facts and Models*. Berlim-Nova York: Springer.

ZATORRE, R. J. (1997). Cerebral Correlates of Human Auditory Processing: Perception of Speech and Musical sounds. In SYKA, J. (*org.*). *Acoustical Signal Processing in the Central Auditory System*. [s.l.], Plenum Press, p. 453-468.



Scanned with CamScanner

O SÍMBOLO CONVIDA A EXPLORAR EM NOSSO SITE O TEMA EM ESTUDO

https://www.flipsnack.com/parabolaeditorial/prosodia2-fxul43fnv/full-view.html

Uma passagem desta *Prosódia*, de Plínio A. Barbosa, resume de maneira exemplar o conteúdo da obra: "A prosódia é o componente de nossa fala que organiza nossos enunciados, moldando nossa maneira de falar através do concurso de modificações articulatórias que se manifestam acusticamente em unidades prosódicas. Essas unidades organizam a nossa fala em níveis que vão da sílaba ao enunciado entoacional".

Plínio A. Barbosa expõe ao leitor, de maneira simples e com a precisão e solidez teórica que o caracterizam, a estruturação dos níveis prosódicos da fala, abordando os aspectos perceptivos, articulatórios e funcionais concernentes. São contempladas as características da entoação, do ritmo, da qualidade de voz, da sílaba, da taxa de elocução, da taxa de articulação e da pausa. A conceituação dos elementos prosódicos, os exemplos e as ilustrações com gráficos cativam e esclarecem o leitor ávido por conhecimentos sobre um aspecto da fala que apresenta alta relevância comunicativa.

*Prosódia*, da nova coleção da Parábola Editorial, "Linguística para o Ensino Superior", é obra que emerge para permanecer como referência intemporal, fruto amadurecido de pesquisa experimental sólida.

Sandra **MADUREIRA**, PUC-SP

ISBN 978-85-7934-163-2
9 788579 341632

Scanned with CamScanner